MULTISCALE DISCRETIZATION OF ELECTRIC-FIELD EQUATIONS

By

SHU-JEN HUANG

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2005

Copyright 2005

by

Shu-Jen Huang

To Deng-Shan,

for now and forever.

ACKNOWLEDGMENTS

First of all, I would like to express my gratitude to my supervisory committee chair, Professor William W. Hager. Without his assistance, this dissertation could not have been completed. In addition, I appreciate his encouragement throughout this period of the research. His confidence in me always strengthens my beliefs in my ability to complete this research.

Second, I would also like to thank Dr. Jay Gopalakrishnan, Dr. Shari Moskow, Dr. Sergei S. Pilyugin, and Dr. Vladimir A. Rakov for serving on my supervisory committee. Their valuable suggestions have been very helpful to my research.

Third, thanks go to my officemates (Dr. Soonchul Park, Beyza, Hongchao, and Sukanya), and all colleagues and friends in the Department of Mathematics at the University of Florida. Their company alleviated the stress and frustration of this time.

Last, but not least, I wish to express my special thanks to my family: to my husband, Deng-Shan, for his love, patience, and encouragement; to our daughter, Marie, for being a glorious joy to us; to my mother for her support and assistance and for helping me to take care of my daughter throughout the dissertation-writing process; to my parents-in-law for their wholehearted understanding and encouragement during my studies; and to my sister for her unstopping support and encouragement. They have always stood by me. Without their support, this dissertation could not have been completed successfully.

TABLE OF CONTENTS

page

ACKNOWLEDGMENTS iv

LIST OF TABLES vii

LIST OF FIGURES viii

KEY TO ABBREVIATIONS AND SYMBOLS x

ABSTRACT xiii

CHAPTER

1 INTRODUCTION 1

2 LITERATURE REVIEW 4

2.1 Dynamic Cloud Models 4
2.2 Approaches for Lowering Charge due to Lightning 6
2.2.1 Lightning Model with Bulk Flashes 6
2.2.2 Lightning Model with Explicit Lightning Channels 7
2.2.3 Hager's Model 10

3 THE MODEL 12

3.1 Formulation of the Equations 12
3.2 Discretized Equations 14
3.3 Important Role and Approximation for Conductivity Function . . 22
3.4 Discussion: $d = \gamma_z(1 - e^{-\gamma(\Delta t)})/\gamma$ 26

4 STABILITY AND ERROR ANALYSIS 28

4.1 Stability 28
4.2 Error Analysis 31

5 CONVERGENCE ANALYSIS 38

5.1 Max-Norm Bound for Matrix $\mathbf{C}^{-1} = \left(\mathbf{A} - \frac{h}{2}\mu\mathbf{B}\right)^{-1}$ 38
5.2 Estimates for $\mathbf{A_1}^{-1}$ 40
5.3 Max-Norm Bound for Matrix $\mathbf{C}^{-1}\mathbf{D}$ 46
5.3.1 Max-Norm for Matrix $\mathbf{C}^{-1}\mathbf{D}$ in 1-Dimensional Space 47
5.3.2 Max-Norm for Matrix $\mathbf{C}^{-1}\mathbf{D}$ in 2- and 3-Dimensional Spaces 52
5.3.3 Numerical Results for Spectral Radius of $\mathbf{C}^{-1}\mathbf{D}$ 57

6 NUMERICAL EXPERIMENTS 59

6.1 Experiment 1 61
6.2 Experiment 2 62
6.3 Experiment 3 66
6.4 Experiment 4 69

7 CONCLUSION 74

APPENDIX 75

A APPROXIMATION FOR CONDUCTIVITY FUNCTION 75

A.1 Approximation from Figure 9.6 [27] 75
A.2 Approximation from Figure 3–3 75

B FISHPAK 79

REFERENCES 80

BIOGRAPHICAL SKETCH 83

LIST OF TABLES

Table		page
3–1	Coefficients of the polynomial $p(z)$	25
5–1	Spectral radius for $\mathbf{C}^{-1}\mathbf{D}$ in one dimensional space	57
5–2	Spectral radius for $\mathbf{C}^{-1}\mathbf{D}$ in two dimensional space	58
5–3	Spectral radius for $\mathbf{C}^{-1}\mathbf{D}$ in three dimensional space	58
6–1	Breakdown time comparison for different values of ω	62
6–2	Relative error in the potential at $t = 65$ seconds using the maximum norm for Crank-Nicholson scheme ($\mu = 1/2$).	67
6–3	Relative error in the potential at $t = 65$ seconds using the maximum norm for Backward Euler scheme ($\mu = 1$).	67
6–4	Estimates for the constant C generated by Equation (6.7)	71
6–5	Estimates for constants p and C generated by Equation (6.10)	73
A–1	Four points chosen from Figure 9.6 [27]	76
A–2	Thirteen points chosen from Figure 3–3	77

LIST OF FIGURES

Figure page

3–1 Path $P_i\vec{P}_j$ connects the node P_i to the node P_j 19

3–2 Directed graph, $G(\mathbf{M})$, for a tridiagonal matrix $\mathbf{M}$ with nonzero entries on its diagonal, superdiagonal, and subdiagonal. 21

3–3 Electrical conductivity σ and corresponding relaxation time $\tau = \varepsilon\sigma^{-1}$, where $\varepsilon = 8.85^{-12}$ F m^{-1}, versus altitude under a variety of geophysical conditions. LL, low latitude, "wavy"; MLPS, mid-latitude typical night (high-latitude, quite); AZTDN, auroral zone, typical disturbed night; MLD, mid-latitude day, quite; MHL, mid-high-latitude, typical of $\sim$ 100 measurements; REP, relativistic electron (energy from a few MeV to 10 MeV) precipitation event (unusual); PCA, polar cap absorption event (an unusually large flux of energetic, $\sim$ 100 MeV) solar protons within the polar cap). Adapted from Hale [8]. (Source: reprinted with permission from Rakov and Uman [21], published in 2003, copyright Vladimir A Rakov and Martin A Uman, published by Cambridge University Press) 24

4–1 Graph of $f(x) = \dfrac{\sin^2 x}{(a+\cos x)^2}$, when $a = -1.05$. 30

4–2 Graph of $f(x) = \dfrac{\sin^2 x}{(a+\cos x)^2}$, when $a = -1$. 31

6–1 Cylindrical shell of the volume integral $\displaystyle\int dV$ 60

6–2 Potentials ($\phi(0,0,z), 0 \le z \le 100$ km) in two domains, Ω_1 and Ω_2 . . 64

6–3 Potentials ($\phi(0,0,z)$, 4.6 km $\le z \le$ 5.4 km) near the negative charge (5 km) in two domains, Ω_1 and Ω_2 64

6–4 Potentials ($\phi(0,0,z)$, 9.6 km $\le z \le$ 10.4 km) near the positive charge (10 km) in two domains, Ω_1 and Ω_2 65

6–5 Potentials ($\phi(x,0,5)$ and $\phi(x,0,10)$, -25 km $\le x \le$ 25 km) along the $x-$axis at the altitude 5 km and 10 km 65

6–6 Maximum norm of potential versus time, where $\Delta t = 1/8$ s and $\mu = 0$ 68

A–1 Electric conductivity as a function of altitude approximated from curve B in Figure 9.6 [27] 76

A–2 Electric conductivity as a function of altitude approximated from Figure 3–3 78

KEY TO ABBREVIATIONS AND SYMBOLS

The list shown below gives a brief description of the major mathematical symbols defined in this dissertation.

$\mathbf{E}$ electric field

$\mathbf{J}$ current density

σ conductivity

ε permittivity

ϕ potential

$$\gamma = \frac{\sigma}{\varepsilon}$$

$$\gamma_z = \frac{\sigma_z}{\varepsilon}$$

$$\lambda = e^{-\gamma(\Delta t)}$$

$$d = \frac{1 - e^{-\gamma(\Delta t)}}{\gamma}\gamma_z$$

$$= \left(1 - e^{-\gamma(\Delta t)}\right)\frac{\sigma_z}{\sigma}$$

$\Omega = \{(x, y, z) : 0 \leq z \leq D_z, |x| \leq D_x, |y| \leq D_y\}$,
where D_x, D_y, and D_z are some prescribed values.

$\Omega_1 = \{(x, y, z) : |x| \leq 50 \text{ km}, |y| \leq 50 \text{ km}, 0 \leq z \leq 100 \text{ km}\}$

$\Omega_2 = \{(x, y, z) : |x| \leq 25 \text{ km}, |y| \leq 25 \text{ km}, 0 \leq z \leq 100 \text{ km}\}$

$$\mathbf{A_1} = \begin{pmatrix} 2 & -1 & & \\ -1 & 2 & -1 & \\ & \ddots & \ddots & \ddots \end{pmatrix}_{m\times m}$$

$$\mathbf{B_1} = \begin{pmatrix} 0 & d_1 & & & \\ -d_2 & 0 & d_2 & & \\ & -d_3 & 0 & d_3 & \\ & & \ddots & \ddots & \ddots \end{pmatrix}_{m\times m}$$

$$\mathbf{\Lambda_1} = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m \end{pmatrix}_{m\times m}$$

$$\mathbf{A_2} = \begin{pmatrix} \mathbf{A_1}+\mathbf{I} & -\mathbf{I} & & \\ -\mathbf{I} & \mathbf{A_1}+\mathbf{I} & -\mathbf{I} & \\ & \ddots & \ddots & \ddots \end{pmatrix}_{m^2\times m^2}$$

$$\mathbf{B_2} = \begin{pmatrix} \mathbf{B_1} & & \\ & \mathbf{B_1} & \\ & & \ddots \end{pmatrix}_{m^2\times m^2}$$

$$\mathbf{\Lambda_2} = \begin{pmatrix} \mathbf{\Lambda_1} & & \\ & \mathbf{\Lambda_1} & \\ & & \ddots \end{pmatrix}_{m^2\times m^2}$$

$$\mathbf{A} = \frac{1}{h^2}\begin{pmatrix} \mathbf{A_2}+\mathbf{I} & -\mathbf{I} & & \\ -\mathbf{I} & \mathbf{A_2}+\mathbf{I} & -\mathbf{I} & \\ & \ddots & \ddots & \ddots \end{pmatrix}_{m^3\times m^3}$$

$$\mathbf{B} = \frac{1}{h^2}\begin{pmatrix} \mathbf{B_2} & & \\ & \mathbf{B_2} & \\ & & \ddots \end{pmatrix}_{m^3\times m^3}$$

$$\mathbf{\Lambda} = \begin{pmatrix} \mathbf{\Lambda_2} & & \\ & \mathbf{\Lambda_2} & \\ & & \ddots \end{pmatrix}_{m^3\times m^3}$$

$$\mathbf{C} = \mathbf{A} - \frac{h}{2}\mu\mathbf{B}$$

$$\mathbf{D} = \mathbf{\Lambda A} + \frac{h}{2}(1-\mu)\mathbf{B}$$

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

MULTISCALE DISCRETIZATION OF
ELECTRIC-FIELD EQUATIONS

By

Shu-Jen Huang

May 2005

Chair: William W. Hager
Major Department: Mathematics

We develop stable finite-difference approximations for equations describing the electric field between the surface of the Earth and the ionosphere. In this scheme, we introduce a parameter μ in the time discretization. For example, $\mu = 0$ and $\mu = 1$ give us forward and backward difference methods respectively, and $\mu = 1/2$ is the Crank-Nicholson scheme. The approximations are unconditionally stable when $1/2 \leq \mu \leq 1$ and unstable when $0 \leq \mu < 1/2$. Max-norm error estimates for the discrete approximations are established. We show that the error is $O(\Delta t)^2 + O(h^2)$ if $\mu = 1/2$, and $O(\Delta t) + O(h^2)$ otherwise. We conduct numerical experiments to verify our analysis.

CHAPTER 1
INTRODUCTION

Most people are afraid of lightning because it can destroy buildings and numerous systems critical to daily life. Moreover, it can kill people. Loss of US property by lightning damage is a huge expense each year.

The physical processes involved in lightning are the focus of intensive research. As the particles, such as graupel and ice, within a cloud grow in size and increase in number, some become charged through collisions. Smaller particles tend to acquire positive charge, while larger particles acquire more negative charge. These particles tend to separate under the influence of updrafts and gravity, until the upper portion of the cloud obtains a net positive charge and the lower portion of the cloud becomes negatively charged. This separation of charge produces huge electrical potential difference within the cloud as well as between the cloud and the ground. As a result of these electric fields, a flash occurs, moving charge between positive and negative regions of a thunderstorm.

Uman [26] gave a detailed history of early lightning research. In the mid eighteenth century, Benjamin Franklin performed the first scientific study of lightning. He designed an experiment that conclusively proved the electric nature of lightning. Little significant progress was made in understanding the properties of lightning until the late nineteenth century when photographic and spectroscopic tools became available for lightning research. Lightning current measurements were first proposed by Pockels [18, 19, 20]. He analyzed the magnetic field induced by lightning currents to estimate the amount of current. Wilson [31, 32] was the first researcher to use the electric field measurements to estimate the structure of thunderstorm charges involved in lightning discharges. He won the Nobel Prize

for inventing the Cloud Chamber and made a major contribution to our present understanding of lightning. Lightning research continued at a steady pace until the late 1960s when the research became particularly active. This increased interest was motivated by

- The danger of lightning to aircraft or spacecraft,
- Solid-state electronics used in computers and other devices,
- Improved measurement and observational capabilities due to advancing technology.

Most lightning research is done by physicists, chemists, meteorologists, and electrical engineers. Hager [5, 6, 7] was the first mathematician using Maxwell's equations to develop a three-dimensional mathematical electrical model to simulate a lightning discharge. His discharge model [6] was obtained by discretizing Maxwell's equations to obtain a relation between the potential field and current density due to the motion of charged particles moving under the influence of wind. Spatial derivatives in his equation were approximated by using volume elements in space, while the temporal derivatives were estimated by a backward Euler scheme in time. Since conductivity is very large in the region where the electric field reaches the breakdown threshold, he evaluated the solution limit as the conductivity tends to infinity in the breakdown region. In his model [6], the output was the electric field as a function of time, and the inputs were currents generated by the flow of charged particles within the thundercloud under the influence of wind.

This dissertation is based on Hager's mathematical model. Some improvements are made compared to Hager's earlier work. For example,

- The equations are discretized in a different way allowing us to solve them by fast fourier transform (FFT) rather than by Cholesky factorization. Consequently, finer meshes can be employed and larger domain can be modeled. For example, numerical results reported by Hager et al. [7], the top

of the model domain was 48 kilometers, we measure the top of the domain at 100 kilometers (ionosphere).

- The huge increase in conductivity with altitude is taken in account in our scheme.

CHAPTER 2
LITERATURE REVIEW

Numerous studies in lightning from different aspects have been reported in the past few decades. This review is focused on numerical models of lightning discharges.

2.1 Dynamic Cloud Models

Atmospheric scientists, Helsdon and Farley [10] and Ziegler et al. [33, 34, 35], considered the electromicrophysical processes for the charge exchange between interacting particles within the cloud domain to obtain dynamic cloud models. They divided water and ice hydrometeors into five categories: cloud water, cloud ice, rain, snow, and high-density precipitating ice (graupel/hail). Each of the five categories of hydrometeor was allowed to have a charge associated with it. This charge was then transported with the species of hydrometeor, according to the transport equation

$$\frac{\partial Q}{\partial t} = -\mathbf{V} \cdot \nabla Q + \frac{1}{\rho_a}\frac{\partial}{\partial z}(U_t Q \rho_a) + \nabla \cdot K_m \nabla Q \pm \left(\frac{\delta Q}{\delta t}\right)_{\text{inter}}, \tag{2.1}$$

where Q was the charge density (in coulombs per cubic meter) carried by hydrometeor category (negative or positive), $\mathbf{V}$ was the velocity vector, ρ_a was the air density, U_t was the mass-weighted mean terminal fall speed of the hydrometeor class, K_m was the nonlinear eddy coefficient, and $(\delta Q/\delta t)_{\text{inter}}$ was the charge exchanged between categories of hydrometeor due to interactions. The cloud model of Helsdon and Farley [10] was two-dimensional; while the models of Ziegler, Straka, MacGorman, Dye, and Ray [33, 34, 35] were established in three-dimensional space. The first term on the right-hand side of Equation (2.1) was the advection term. The second term was the fallout term, and it was equal to zero for cloud-water

and cloud-ice charge. The third term was the eddy-mixing term, and the last term represented the charge exchanged between any two classes of hydrometeor during an interaction between those two classes.

Helsdon and Farley's model [10] also considered the presence of small ions. The growth and decay equation governing the number concentration of small ions was

$$\frac{\partial n_{1,2}}{\partial t} = -\nabla \cdot (n_{1,2}\mathbf{V} \pm n_{1,2}\mu_{1,2}\mathbf{E} - K_m \nabla n_{1,2}) + G - \alpha n_1 n_2 + \mathrm{SRC} - \mathrm{SINK},$$

where $n_{1,2}$ was the concentration of small ions with subscript 1 representing positive ions and subscript 2 representing negative ions, μ was the small-ion mobility (in meters squared per volt per second, pressure dependent), $\mathbf{E}$ was the two-dimensional electric field vector, G was the ion-generation rate due to cosmic ray ionization (in ions per cubic meter per second, dependent on height), α was the ion-ion recombination coefficient (1.6×10^{-12} m^3/s), SRC was the source term for small ions from processes other than cosmic-ray ionization, and SINK was the sink term for small ions from processes other than recombination. Ziegler et al. [34] excluded the contribution from free ions in their model, because they thought this contribution was very small within the cloud.

With the various types of charges accounted for in the model, at each grid point a total charge density can be expressed as

$$\rho = e(n_1 - n_2) + \sum Q,$$

where ρ was the total charge density (in coulombs per cubic meter) and e was the electronic charge (1.6×10^{-19} C). The first term represented the net charge due to the difference in small-ion concentrations, and it was equal to zero in the model of Ziegler et al. [34] because they ignored the contribution from free ions. The second term represented the net charge on the five categories of hydrometeor.

With the total charge determined, they applied Poisson's equation

$$\nabla^2 \phi = \frac{\rho}{\varepsilon}$$

to obtain the scalar electric potential ϕ in volts, where ε was the electrical permittivity of air (8.8592×10^{-12} C^2 / (N m^2)). Then the electric-field vector can be solved from the potential ϕ using Gauss's law

$$\mathbf{E} = -\nabla \phi.$$

2.2 Approaches for Lowering Charge due to Lightning

There are different approaches for lowering charge due to lightning. The neutralization of charge by lightning in the models with bulk flashes and with explicit lightning channels are discussed in Section 2.2.1 and Section 2.2.2, respectively. The approach used in the model of Hager et al. is discussed in Section 2.2.3.

2.2.1 Lightning Model with Bulk Flashes

Rawlins [22] was the first scientist to add a simple lightning parameterization to a numerical cloud model. When the electric field reached the discharge threshold of 500 kV/m everywhere within the model domain, he simply reduced the net charge by a constant fraction everywhere (arbitrarily to 30% of the pre-strike values). This was the simplest representation of the neutralization of charge by lightning.

On the other hand, Takahashi [25] suggested a more complex parameterization of a lightning flash for his axisymmetric, two-dimensional model. His parameterization neutralized 20 Coulombs of charge by removing that amount in each of the two oppositely charged regions whenever the maximum electric field had the value of 340 kV/m. To neutralize the charge, first he found the maximum positive charge density and the maximum negative charge density. Next, he determined the smallest region about each of the two maxima that contained 20 Coulombs. Finally, he

subtracted positive charge equally from all charged particles in the positive region; and similarly, subtracted negative charge equally from all charged particles in the negative region.

Ziegler and MacGorman [33] also applied a simple parameterization, but used a three-dimensional model to treat the bulk effect of lightning on storm charge in a time step. They neutralized a fraction of the net charge at all grid points where the magnitude of the net charge density exceeded 0.5 nC/m^3 to imitate the effect of lightning releasing the charge from its channels. The above bulk-flash schemes are relatively easy to implement but lack realism.

2.2.2 Lightning Model with Explicit Lightning Channels

Most lightning discharge models with explicit channel treat only a one-dimensional or unbranched channel. The branching of lightning channels would be very difficult to treat analytically because of the lack of symmetry and the incomplete understanding of the processes.

2.2.2.1 The Model of Helsdon et al.

Helsdon et al. [9, 10, 11, 12] estimated both the geometry and charge distribution of an intercloud lightning flash in a two-dimensional Storm Electrification Model (SEM). Since then, their parameterization has been extended to a three-dimensional numerical cloud model. They applied the concept of bidirectional leader [13]: The parameterized lightning propagated bidirectionally (initially parallel and antiparallel to the electric field) from the point of initial breakdown.

Their parameterization produced a single, unbranched channel that traced the electric-field line from an initial point. Initiation, propagation direction, and termination of the discharge were computed using the magnitude and direction of the electric field vector as the determining criteria. The charge redistribution associated with lightning was approximated by assuming that the channel remained electrically neutral over its total length.

An initial criterion: a threshold of electric field was chosen with a value of 400 kV/m. The channel was extended in both directions along the field line until the ambient electric-field magnitude fell below a certain threshold (150 kV/m) at the locations of the channel-termination points. They assumed that the linear charge density at a grid point, P, along the channel was proportional to the difference between the potential at the point where the discharge originated, and the potential at P. The linear charge density can be presented by

$$Q_P = -k(\Phi_P - \Phi_0),$$

where Q_P was the charge density at P, and Φ_P and Φ_0 were the potentials at P and the initiation point of the discharge, respectively. The value of this proportionality constant k controlled the amount of charge transferred by the discharge. Charge distribution at each end of the channel was adjusted in order to maintain net charge neutrality of the flash, consistent with the theory proposed by Kasemir [13]. They extended the channel by four grid points at each end, and adjusted the distribution of linear charge density on the extensions, to force the integral of lightning charge density for the whole channel to be zero. In this extended region, they assumed that the charge density decreased like $e^{-\alpha x^2}$, where x is the distance from the channel.

2.2.2.2 MacGorman's Model

MacGorman et al. [15] suggested a lightning parameterization that was considered an extension of the Helsdon et al. [12] parameterization in conjunction with some of the bulk-lightning parameterization methods presented by Ziegler and MacGorman [33]. MacGorman et al. [15] developed a parameterization to enable cloud models to simulate the location and structure of individual lightning flashes by using the conceptual model of MacGorman et al. [14] and Williams et al. [30]. Their parameterization [15] of lightning flash structure proceeded in two stages.

First, Helsdon et al. [12] provided the initial development. From initiation at a grid point, a flash traced the electric-field line outward in both parallel and antiparallel directions, until the magnitude of the ambient electric field at each end fell below some certain threshold value. Secondly, if one end of the channel reached ground, the parameterization terminated at that end, but allowed the other end to continue developing.

Charge estimation and neutralization were parameterized by applying the technique proposed by Ziegler and MacGorman [33], except that Ziegler and MacGorman neutralized charge at all grid points having $|\rho(i,j,k)| \geq \rho_1$ (where $\rho(i,j,k)$ was the net charge density at the grid point (i,j,k) and ρ_1 was the minimum $|\rho(i,j,k)|$ for all grid points to be involved in lightning beyond initial propagation) throughout the storm, but their parameterization neutralized charge only at such grid point within a single localized flash.

2.2.2.3 Mansell's Model

Mansell et al . [16] proposed a lightning parameterization derived from the dielectric breakdown model that was developed by Niemeyer et al. [17] and Wiesmann and Zeller [29] to simulate electric discharges. They extended the dielectric breakdown model to a three-dimensional domain to represent more realistic electric field and thunderstorm dynamics.

In their work, the stochastic lightning model (SLM) was an application of the Wiesmann-Zeller model to simulate bidirectional discharges in the regions of varying net charge density (e.g., in an electrified thunderstorm). Procedures for simulating lightning flashes in the thunderstorm model were as follows. A flash occurred when the magnitude of the electric field exceeded the initiation threshold E_{init} anywhere in the model domain. The lightning initiation point was chosen randomly from among all the points where the magnitude of the electric field is greater than $0.9E_{init}$. Both decisions for choosing the initiation threshold

and the initiation point were suggested by MacGorman et al. [15]. Positive and negative parts of the flash were propagated independently so that up to two new channel segments (positive and negative) could be added at each step. Both ends had default initial propagation thresholds of $0.75E_{init}$. For flash neutrality, they assumed that the channel structure would maintain overall charge neutrality as long as neither end reached the ground, as suggested by Kasemir [13]. But, for computational simplicity, their parameterization maintained near-neutrality (within 5%) by a technique of adjusting the reference potential to the growth of the lightning structure instead of adjusting the reference potential of the channel.

2.2.3 Hager's Model

Hager et al. [5, 6, 7] proposed a three-dimensional lightning-discharge model that produced bidirectional, axisymmetric IC and -CG flashes. The model generated the discharge region, charge transfer, and detailed charge rearrangement associated with the flash.

Their approach to lightning was quite different from those in Section 2.2.1 and Section 2.2.2. Their breakdown model was based on Maxwell's equations. They assumed that current due to transport of charge under the influence of wind was known. They obtained an equation governing the evolution of the electric potential under the assumption that the time derivative of the magnetic field can be disregarded. After integrating this equation over boxes and approximating derivatives by finite differences, they obtained an implicit system of difference equations describing the evolution of the electric field. Their approach to lightning was to let the conductivity tend to infinity wherever the electric field reached the breakdown threshold. This approach appeals to our basic conception of nature: When the electric field reaches breakdown threshold, conductivity becomes very large as a plasma forms.

When the electric field reaches the breakdown threshold, the electric potential changes instantaneously everywhere within the thundercloud. The Inverse Matrix Modification Formula [4] was applied to evaluate this change:

$$\mathbf{\Phi}^{\text{after}} = \mathbf{\Phi}^{\text{before}} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{U}^{\top}\mathbf{A}^{-1}\mathbf{U})^{-1}\mathbf{U}^{\top}\mathbf{\Phi}^{\text{before}}, \tag{2.2}$$

where $\mathbf{\Phi}^{\text{before}}$ was the electric potential before discharge, $\mathbf{\Phi}^{\text{after}}$ was the electric potential after discharge, $\mathbf{A}$ was the discrete Laplacian, and $\mathbf{U}$ was a matrix with a +1 and -1 in each column corresponding to the arcs associated with the breakdown. There were no parameters in Equation (2.2) besides the electric potential before discharge. This was consistent with experimental observations: The charge is controlled predominately by a single parameter: the local electrostatic field. This was observed in experiments reported by Williams et al. [30].

CHAPTER 3
THE MODEL

Hager et al. [6] were the first to use Maxwell's equations to develop a three-dimensional electric model for a thunderstorm. We first establish the model based on Maxwell's equations.

3.1 Formulation of the Equations

According to Hager et al. [6], the curl of the magnetic field strength $\mathbf{H}$ is given based on Maxwell's equations:

$$\nabla \times \mathbf{H} = \varepsilon \frac{\partial \mathbf{E}}{\partial t} + \sigma \mathbf{E} + \mathbf{J}, \tag{3.1}$$

where ε is the permittivity, σ is the conductivity, $\mathbf{E}$ is the electric field, and $\mathbf{J}$ is the current density associated with the influence of the wind. In this model, we have the following assumptions:

- Assumption 1: The time derivative of the magnetic field is zero, i.e.,

$$\frac{\partial \mathbf{B}}{\partial t} = \mathbf{0} \;\Rightarrow\; \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} = \mathbf{0}$$
$$\Rightarrow\; \mathbf{E} = \nabla \phi$$

  Therefore $\mathbf{E}$ is the gradient of a potential ϕ.

- Assumption 2: Let $\mathbf{E}_B$ be the breakdown field strength. Then the electric field magnitude is always less than or equal to the breakdown threshold $\mathbf{E}_B$. That is, $|\mathbf{E}| \leq \mathbf{E}_B$.

- Assumption 3: When the electric field reaches the breakdown threshold $\mathbf{E}_B$ at some point, the conductivity tends to infinity in a small neighborhood at that point. That is, if $|\mathbf{E}(x, y, z)| = \mathbf{E}_B$, then $\sigma(x, y, z) = +\infty$.

In this dissertation, we do not implement the lightning discharge described in Section 2.2.3. Instead we focus on integrating the equations up to the discharge initiation.

We take the divergence of Equation (3.1) to get

$$\varepsilon \nabla \cdot \frac{\partial \mathbf{E}}{\partial t} + \nabla \cdot \sigma \mathbf{E} + \nabla \cdot \mathbf{J} = 0. \tag{3.2}$$

Boundary conditions: In this model, we consider a bounded domain Ω between the surface of the earth and ionosphere. Let us consider boundary conditions on $\partial\Omega$:

- $\phi = 0$ on the surface of the earth, since we treat the earth as a perfect conductor.
- $\phi = \phi_1$ on the ionosphere. $\phi_1 = 250{,}000$ V (Adlerman and Williams [1]).
- On the sides of the domain Ω, assume that the potential ϕ is the "fair field" potential. That is, the current and the time derivative of the electric field are both zero under fair-weather conditions. Hence, the potential ϕ satisfies

$$\nabla \cdot \sigma \mathbf{E} = 0. \tag{3.3}$$

Assume that both the potential ϕ and the conductivity σ depend only on z on the sides of the domain Ω, and since $\mathbf{E} = \nabla\phi$, we have

$$\frac{\partial}{\partial z}\left(\sigma \frac{\partial}{\partial z}\phi\right) = 0, \tag{3.4}$$

from Equation (3.3). In order to obtain the potential function on the sides of the domain Ω, the Equation (3.4) is solved with these boundary conditions: the potential is zero on the surface of Earth and ϕ_1 on the ionosphere. Therefore, we discover that

$$\bar{\phi}(z) = \int_0^z \frac{c}{\sigma(z)} dz,$$

where

$$c = \frac{\phi_1}{\displaystyle\int_0^h \frac{1}{\phi(z)} dz},$$

and h is the height of the domain Ω ($h = 100$ km for numerical experiments in Chapter 6). If $\phi_1 = 250000$ (see [1]), $c = 2.852108 \times 10^{-9}$.

- Finally, using the transformation $\phi(x, y, z) = \bar{\phi}(z) + \psi(x, y, z)$, we have the boundary condition: $\psi = 0$ on $\partial\Omega$.

Since $\mathbf{E} = \nabla\phi$ and $\phi = \bar{\phi} + \psi$, Equation (3.2) becomes

$$\varepsilon \frac{\partial}{\partial t} \triangle\psi + \sigma \triangle\psi + \sigma_z \frac{\partial \psi}{\partial z} + \nabla \cdot \mathbf{J} = 0,$$

where $\triangle$ denotes the Laplacian operator defined by $\triangle = \nabla \cdot \nabla$, and σ_z denotes the derivative of σ with respect to z.

Then our model is the partial differential equation with the boundary condition as follows:

$$\begin{aligned} \text{PDE:} \quad & \frac{\partial}{\partial t} \triangle\psi + \frac{\sigma}{\varepsilon} \triangle\psi + \left(\frac{\sigma_z}{\varepsilon} \frac{\partial \psi}{\partial z} + \frac{\nabla \cdot \mathbf{J}}{\varepsilon} \right) = 0 \\ \text{BC:} \quad & \psi = 0 \text{ on } \partial\Omega \\ \text{Domain:} \quad & \Omega = \{(x, y, z) : 0 \le z \le D_z, |x| \le D_x, |y| \le D_y\}, \end{aligned}$$

where D_x, D_y, and D_z are prescribed values. In our computation, we take $D_z = 100$ km. For reference,

$$\Omega_1 = \{(x, y, z) : |x| \le 50 \text{ km}, |y| \le 50 \text{ km}, 0 \le z \le 100 \text{ km}\}. \tag{3.5}$$

3.2 Discretized Equations

We demonstrate the discretization process for a uniform mesh in a rectangular coordinate system. Consider the differential equation

$$\frac{\partial}{\partial t} \triangle\psi = -\frac{\sigma}{\varepsilon} \triangle\psi - \left(\frac{\sigma_z}{\varepsilon} \frac{\partial \psi}{\partial z} + \frac{\nabla \cdot \mathbf{J}}{\varepsilon} \right). \tag{3.6}$$

The solution of Equation (3.6) is

$$\triangle\psi^{n+1} = e^{-\gamma(\triangle t)}\triangle\psi^{n} - \int_{t_n}^{t_{n+1}} e^{\gamma(s-t_{n+1})}\left(\gamma_z\frac{\partial\psi}{\partial z} + \frac{\nabla\cdot\mathbf{J}}{\varepsilon}\right)ds, \tag{3.7}$$

where superscript n denotes the time level t_n, $\gamma = \sigma/\varepsilon$, and $\gamma_z = \sigma_z/\varepsilon$ is the derivative of σ/ε with respect to z. First, we approximate the integral part of Equation (3.7), denoted as I:

$$I = \int_{t_n}^{t_{n+1}} e^{\gamma(s-t_{n+1})}\left(\gamma_z\frac{\partial\psi}{\partial z} + \frac{\nabla\cdot\mathbf{J}}{\varepsilon}\right)ds.$$

The following steps are conducted.

- Step 1: Both $\partial\psi/\partial z$ and $\nabla\cdot\mathbf{J}$ are evaluated at time $\frac{1}{2}(t_n + t_{n+1})$, which is denoted as time level $n+\frac{1}{2}$. Hence we have the following approximation

$$I \approx \frac{1-e^{-\gamma(\triangle t)}}{\gamma}\left(\gamma_z\left(\frac{\partial\psi}{\partial z}\right)^{n+\frac{1}{2}} + \frac{(\nabla\cdot\mathbf{J})^{n+\frac{1}{2}}}{\varepsilon}\right).$$

  Define

$$d = \frac{1-e^{-\gamma(\triangle t)}}{\gamma}\gamma_z.$$

  Then we can rewrite the above approximation as follows :

$$I \approx d\left(\frac{\partial\psi}{\partial z}\right)^{n+\frac{1}{2}} + \frac{1-e^{-\gamma(\triangle t)}}{\gamma}\frac{(\nabla\cdot\mathbf{J})^{n+\frac{1}{2}}}{\varepsilon}.$$

- Step 2: Use $(1-\mu)(\partial\psi/\partial z)^n + \mu(\partial\psi/\partial z)^{n+1}$ to approximate the term $(\partial\psi/\partial z)^{n+\frac{1}{2}}$. Hence,

$$I \approx d\left[(1-\mu)\left(\frac{\partial\psi}{\partial z}\right)^{n} + \mu\left(\frac{\partial\psi}{\partial z}\right)^{n+1}\right] + \frac{1-e^{-\gamma(\triangle t)}}{\gamma}\frac{(\nabla\cdot\mathbf{J})^{n+\frac{1}{2}}}{\varepsilon},$$

  where $0 \le \mu \le 1$.

So Equation (3.7) is then approximated by the following equation

$$\begin{aligned}\triangle\psi^{n+1} = & \\ & \lambda\triangle\psi^n - d\left[(1-\mu)\left(\frac{\partial\psi}{\partial z}\right)^n + \mu\left(\frac{\partial\psi}{\partial z}\right)^{n+1}\right] \\ & -\frac{1-e^{-\gamma(\triangle t)}}{\gamma}\frac{(\nabla\cdot\mathbf{J})^{n+\frac{1}{2}}}{\varepsilon},\end{aligned} \tag{3.8}$$

where $\lambda = e^{-\gamma(\triangle t)}$, $\gamma = \sigma/\varepsilon$, and $0 \leq u \leq 1$.

Now we consider Equation (3.6) but ignore the force term $\mathbf{J}$.

$$\frac{\partial}{\partial t}\triangle\psi = -\gamma\triangle\psi - \gamma_z\frac{\partial\psi}{\partial z}. \tag{3.9}$$

Similarly, the solution of Equation (3.9) is

$$\triangle\psi^{n+1} = e^{-\gamma(\triangle t)}\triangle\psi^n - \int_{t_n}^{t_{n+1}} e^{\gamma(s-t_{n+1})}\gamma_z\frac{\partial\psi}{\partial z}ds, \tag{3.10}$$

and it can be approximated by

$$\triangle\psi^{n+1} = \lambda\triangle\psi^n - d\left[(1-\mu)\left(\frac{\partial\psi}{\partial z}\right)^n + \mu\left(\frac{\partial\psi}{\partial z}\right)^{n+1}\right] \tag{3.11}$$

Next, the spatial part of Equation (3.11) can be discretized by the finite difference method. Define

$$h_x = \frac{2D_x}{m+1},\ h_y = \frac{2D_y}{m+1},\ \text{and}\ h_z = \frac{D_z}{m+1}.$$

where m is a positive integer. The mesh point (x_i, y_j, z_k) is defined as

$$\begin{aligned} x_i &= -D_x + ih_x, \\ y_j &= -D_y + jh_y, \\ z_k &= kh_z, \end{aligned}$$

where $0 \leq i \leq m+1$, $0 \leq j \leq m+1$, and $0 \leq k \leq m+1$. Let $\psi_{i,j,k}$ denote the discrete approximation to ψ at the point (x_i, y_j, z_k). For domain Ω_1: using the

uniform mesh size for x, y, and z (that is, $\triangle x = \triangle y = \triangle z = h$), we have the approximation

$$\partial_z \psi_{i,j,k} = \frac{\psi_{i,j,k+1} - \psi_{i,j,k-1}}{2h}$$

for the partial derivative $\partial\psi/\partial z$ at the point (x_i, y_j, z_k). The second partial derivative $\partial^2\psi/\partial x^2$ at the point (x_i, y_j, z_k) has an approximation denoted by $\partial_1^2 \psi_{i,j,k}$:

$$\partial_1^2 \psi_{i,j,k} = \frac{\psi_{i+1,j,k} - 2\psi_{i,j,k} + \psi_{i-1,j,k}}{h^2}$$

Similarly, we use approximations $\partial_2^2 \psi_{i,j,k}$ and $\partial_3^2 \psi_{i,j,k}$ to the second partial derivatives $\partial^2\psi/\partial y^2$ and $\partial^2\psi/\partial z^2$:

$$\partial_2^2 \psi_{i,j,k} = \frac{\psi_{i,j+1,k} - 2\psi_{i,j,k} + \psi_{i,j-1,k}}{h^2}$$

and

$$\partial_3^2 \psi_{i,j,k} = \frac{\psi_{i,j,k+1} - 2\psi_{i,j,k} + \psi_{i,j,k-1}}{h^2},$$

respectively. We define the discrete Laplacian operator $\triangle^h$ as follows:

$$\triangle^h \psi_{i,j,k} = \sum_{l=1}^{3} \partial_l^2 \psi_{i,j,k}.$$

After replacing $\triangle$ by $\triangle^h$ and $\partial\psi/\partial z$ by the discrete $\partial_z\psi$ in Equation (3.11), we obtain the following finite difference scheme.

$$-\triangle^h \psi_{i,j,k}^{n+1} + \lambda_k \triangle^h \psi_{i,j,k}^{n} = \frac{h}{2}\left[(1-\mu)\frac{d_k}{h^2}\partial_z \psi_{i,j,k}^{n} + \mu\frac{d_k}{h^2}\partial_z \psi_{i,j,k}^{n+1}\right] \tag{3.12}$$

Equation (3.12) yields a matrix-vector form

$$\mathbf{A}\Psi^{n+1} - \mathbf{\Lambda}\mathbf{A}\Psi^{n} = \frac{h}{2}\left[(1-\mu)\mathbf{B}\Psi^{n} + \mu\mathbf{B}\Psi^{n+1}\right]. \tag{3.13}$$

where Ψ denotes the vector with components $\psi_{i,j,k}$. The matrix $\mathbf{A}$ is symmetric and positive definite and the structure of the matrix $\mathbf{A}$ is a block tridiagonal

matrix in the following form:

$$\mathbf{A} = \frac{1}{h^2}\begin{pmatrix} \mathbf{A_2}+\mathbf{I}_{m^2} & -\mathbf{I}_{m^2} & & & & & \\ -\mathbf{I}_{m^2} & \mathbf{A_2}+\mathbf{I}_{m^2} & -\mathbf{I}_{m^2} & & & & \\ & \ddots & \ddots & \ddots & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & -\mathbf{I}_{m^2} & \mathbf{A_2}+\mathbf{I}_{m^2} & -\mathbf{I}_{m^2} \\ & & & & & -\mathbf{I}_{m^2} & \mathbf{A_2}+\mathbf{I}_{m^2} \end{pmatrix}, \quad (3.14)$$

where $\mathbf{I}_{m^2}$ is the $m^2 \times m^2$ identity matrix and $\mathbf{A_2}$ is an $m^2 \times m^2$ block tridiagonal matrix in the form

$$\begin{pmatrix} \mathbf{A_1}+\mathbf{I}_m & -\mathbf{I}_m & & & \\ -\mathbf{I}_m & \mathbf{A_1}+\mathbf{I}_m & -\mathbf{I}_m & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \ddots \\ & & & \ddots & \ddots \end{pmatrix},$$

where $\mathbf{I}_m$ is an $m \times m$ identity matrix and $\mathbf{A_1}$ is a tridiagonal matrix with 2's on its diagonal and -1's on its superdiagonal and subdiagonal. The matrix $\mathbf{\Lambda}$ is a diagonal matrix with m^2 blocks on the diagonal, and each block has λ_i on its diagonal. The matrix $\mathbf{B}$ is a diagonal block matrix in the following form:

$$\mathbf{B} = \frac{1}{h^2}\begin{pmatrix} \mathbf{B_1} & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \mathbf{B_1} \end{pmatrix}, \quad (3.15)$$

where the diagonal block $\mathbf{B_1}$ is an $m \times m$ tridiagonal matrix in the form

$$\mathbf{B_1} = \begin{pmatrix} 0 & d_1 & & & \\ -d_2 & 0 & d_2 & & \\ & -d_3 & 0 & d_3 & \\ & & \cdot & \cdot & \cdot \\ & & & \cdot & \cdot & \cdot \\ & & & & \cdot & \cdot & \cdot \end{pmatrix}.$$

Varga [28] gives the definition for an irreducible matrix and some theorems to determine the irreducibility of a matrix. Definition 3.2.1, Definition 3.2.2, Theorem 3.2.3, Definition 3.2.4, and Theorem 3.2.5 are adopted from Varga [28].

Definition 3.2.1 *For $n \geq 2$, an $n \times n$ matrix $\mathbf{M}$ is reducible if there exists an $n \times n$ permutation matrix $\mathbf{P}$ such that*

$$\mathbf{PMP}^T = \begin{pmatrix} \mathbf{M}_{1,1} & \mathbf{M}_{1,2} \\ \mathbf{O} & \mathbf{M}_{2,2} \end{pmatrix},$$

where $\mathbf{M}_{1,1}$ is an $r \times r$ submatrix and $\mathbf{M}_{2,2}$ is an $(n-r) \times (n-r)$ submatrix, where $1 \leq r < n$. If no such permutation matrix exists, then $\mathbf{M}$ is irreducible.

A directed graph is an alternative way to check irreducibility for a matrix. Let $\mathbf{M}$ be any $n \times n$ matrix, a finite directed graph $G(\mathbf{M})$ can be constructed in the following way: Consider any n distinct points (nodes) $P_1, P_2, \cdots, P_n$ in the plane. For every nonzero entry $m_{i,j}$ of the matrix, we connect a path directed from the node P_i to the node P_j, as shown in Figure 3–1.

Figure 3–1: Path $P_i\vec{P}_j$ connects the node P_i to the node P_j

Definition 3.2.2 *A directed graph is strongly connected if, for any ordered pair of nodes P_i and P_j, there exists a directed path*

$$P_i\vec{P}_{l_1}\ ,\ P_{l_1}\vec{P}_{l_2}\ ,\cdots,P_{l_{r-1}}\vec{P}_{l_r=j}$$

connecting P_i to P_j.

Theorem 3.2.3 *An $n \times n$ matrix $\mathbf{M}$ is irreducible if and only if its directed graph $G(\mathbf{M})$ is strongly connected.*

Definition 3.2.4 *An $n \times n$ matrix $\mathbf{M} = (m_{i,j})$ is diagonally dominant if*

$$|m_{i,i}| \geq \sum_{j=1, j\neq i}^{n} |m_{i,j}| \tag{3.16}$$

for all $1 \leq i \leq n$. An $n \times n$ matrix $\mathbf{M}$ is irreducibly diagonally dominant if $\mathbf{M}$ is irreducible and diagonally dominant, with strict inequality in Inequality (3.16) for at least one i.

Theorem 3.2.5 *If $\mathbf{M}$ is an $n \times n$ irreducibly diagonally dominant matrix, then the matrix $\mathbf{M}$ is nonsingular.*

Now, we rearrange Equation (3.13) to yield

$$\left(\mathbf{A} - \frac{h}{2}\mu\mathbf{B}\right)\mathbf{\Psi}^{n+1} = \left(\mathbf{\Lambda A} + \frac{h}{2}(1-\mu)\mathbf{B}\right)\mathbf{\Psi}^{n}. \tag{3.17}$$

Let $\mathbf{C} = \mathbf{A} - \frac{h}{2}\mu\mathbf{B}$ and $\mathbf{D} = \mathbf{\Lambda A} + \frac{h}{2}(1-\mu)\mathbf{B}$. Then we have

$$\mathbf{C}\mathbf{\Psi}^{n+1} = \mathbf{D}\mathbf{\Psi}^{n},$$

where the matrix $\mathbf{C}$ is a block tridiagonal matrix given by

$$\mathbf{C} = \frac{1}{h^2}\begin{pmatrix} \mathbf{C}_{m^2} & -\mathbf{I}_{m^2} & & & \\ -\mathbf{I}_{m^2} & \mathbf{C}_{m^2} & -\mathbf{I}_{m^2} & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \ddots \\ & & & \ddots & \ddots \end{pmatrix}, \tag{3.18}$$

where $\mathbf{C}_{m^2}$ is an $m^2 \times m^2$ tridiagonal block matrix with $m \times m$ matrices

$$\begin{pmatrix} 6 & -1-\frac{h}{2}d_1 & & & \\ -1+\frac{h}{2}d_2 & 6 & -1-\frac{h}{2}d_2 & & \\ & \cdot & \cdot & \cdot & \\ & & \cdot & \cdot & \cdot \\ & & & \cdot & \cdot & \cdot \end{pmatrix} \tag{3.19}$$

on the diagonal block, and with negative $m \times m$ identity matrices on both its super-diagonal and sub-diagonal blocks.

Lemma 3.2.6 *Let* $\mathbf{M}$ *be any* $n \times n$ *tridiagonal matrix. If all entries on the diagonal, superdiagonal, and subdiagonal of* $\mathbf{M}$ *are nonzero, then* $\mathbf{M}$ *is irreducible.*

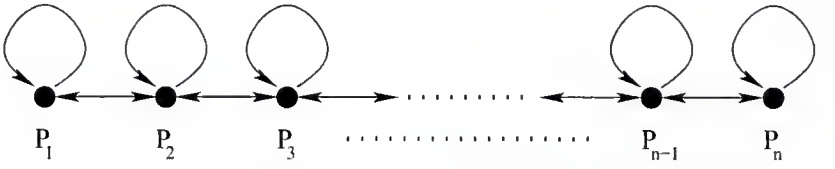

Figure 3–2: Directed graph, $G(\mathbf{M})$, for a tridiagonal matrix $\mathbf{M}$ with nonzero entries on its diagonal, superdiagonal, and subdiagonal.

Proof: Since the tridiagonal matrix $\mathbf{M}$ has nonzero entries on its diagonal, superdiagonal, and subdiagonal, the directed graph of the matrix $\mathbf{M}$, $G(\mathbf{M})$, is shown in Figure 3–2. From Figure 3–2, it is easy to see that $G(\mathbf{M})$ is strongly connected, and therefore $\mathbf{M}$ is irreducible by Theorem 3.2.3. ∎

Lemma 3.2.7 *The matrix* $\mathbf{C}$ *in (3.18) is irreducible.*

Proof: First, we want to show that each diagonal block $\mathbf{C}_{m^2}$ of the matrix $\mathbf{C}$ is irreducible. By Lemma 3.2.6, each diagonal block (3.19) of $\mathbf{C}_{m^2}$ is irreducible; hence, there exists a path for any two vertices associated with a diagonal block (3.19). Suppose we are given vertices $i < j$ in two distinct diagonal blocks; the subdiagonal blocks of $\mathbf{C}_{m^2}$ provides a path between i and $i+m$ and between $i+m$ and $i+2m$ etc. Eventually, we reach a vertex k in the block containing j. And

since the vertices k and j are in the same irreducible diagonal block, there exists a path from k to j. Hence, there exists a path from i to j.

Now the matrix $\mathbf{C}$ has irreducible diagonal blocks $\mathbf{C}_{m^2}$. The structure of $\mathbf{C}$ is similar as the structure of $\mathbf{C}_{m^2}$. Similarly, we can show that the matrix $\mathbf{C}$ is irreducible. ■

Therefore $\mathbf{C} = \mathbf{A} - \frac{h}{2}\mu\mathbf{B}$ is irreducibly diagonally dominant by Definition 3.2.4 and Lemma 3.2.7. Based on Theorem 3.2.5, $\mathbf{C}$ is nonsingular. Therefore, we have the iteration

$$\mathbf{\Psi}^{n+1} = \mathbf{C}^{-1}\mathbf{D}\mathbf{\Psi}^{n},$$

where $\mathbf{C} = \mathbf{A} - \frac{h}{2}\mu\mathbf{B}$, $\mathbf{D} = \mathbf{\Lambda A} + \frac{h}{2}(1-\mu)\mathbf{B}$, and $\mathbf{C^{-1}D}$ is the amplification matrix.

3.3 Important Role and Approximation for Conductivity Function

The conductivity function σ varies by many orders of magnitude between the surface of the Earth and the ionosphere. Figure ?? and Figure 3–3 show that σ grows more than the factor 10^{10} S/m from the surface of the Earth to the ionosphere, which leads the differential Equation (3.6) to be a very stiff equation.

For a stiff equation

$$\frac{dy}{dt} = -ky, \tag{3.20}$$

where k is a very large positive number, its solution is $y(t) = e^{-kt}y(0)$ and $y(t)$ tends to zero quickly. Let us consider the following three schemes:

1. Euler explicit scheme for Equation (3.20):

$$\frac{y^{n+1} - y^n}{\triangle t} = -ky^n.$$

Therefore, $y^n = (1 - k\triangle t)^n y(0)$. When k is very large, $(1 - k\triangle t) < -1$ and $y^n \to \pm\infty$ as n increases. Hence Euler explicit scheme does not work.

2. Euler implicit scheme for Equation (3.20)

$$\frac{y^{n+1} - y^n}{\triangle t} = -ky^{n+1}.$$

Therefore, $y^n = \left(\frac{1}{1+k\triangle t}\right)^n y(0)$ and $0 < \frac{1}{1+k\triangle t} < 1$ when k is large. Hence $y^n \to 0$ as $n \to \infty$. The scheme works.

3. The exact formula:

$$y^{n+1} = e^{-k\triangle t} y^n.$$

Therefore, $y^{n+1} = \left(e^{-k\triangle t}\right)^n y(0)$. The factor $\left(e^{-k\triangle t}\right)^n$ tends to zero much faster than $\left(\frac{1}{1+k\triangle t}\right)^n$ when k is large. This scheme works better than Euler implicit scheme.

Therefore, the above explanation demonstrates that the exact formula is the best method to solve a stiff equation. The conductivity function, σ, plays an important role in the Equation (3.6). That is the reason that the exact solution (3.7) is chosen in Section 3.2.

At the beginning of the research, we use Figure 9.6 in [27] to find an approximation for the conductivity function σ (see Appendix A.1).

- For $0 \le z \le 27.828$ km:

  $\sigma(z) = 10^{-13.40+0.1600z-0.005175z^2+0.00009077z^3}$

- For $z > 27.828$ km:

  $\sigma(z) = 10^{-12.88+0.06718z}$

  Then we have

$$\begin{cases} \dfrac{\sigma_z}{\sigma} = (0.1600 - 0.01035z + 0.0002723z^2)(\ln 10) & \text{, for } 0 \le z \le 27.828 \\ \dfrac{\sigma_z}{\sigma} = 0.1546 & \text{, for } z > 27.828 \end{cases}$$

and

$$\begin{cases} \dfrac{\partial}{\partial z}\left(\dfrac{\sigma_z}{\sigma}\right) = (-0.01035 + 0.0005446z)(\ln 10) & \text{, for } 0 \le z \le 27.828 \\ \dfrac{\partial}{\partial z}\left(\dfrac{\sigma_z}{\sigma}\right) = 0 & \text{, for } z > 27.828 \end{cases}$$

Figure 9.6 [27] only provides conductivity data up to 27.8 km (about 100,000 ft). In order to produce more accurate results, we use Figure 1.3 in the book

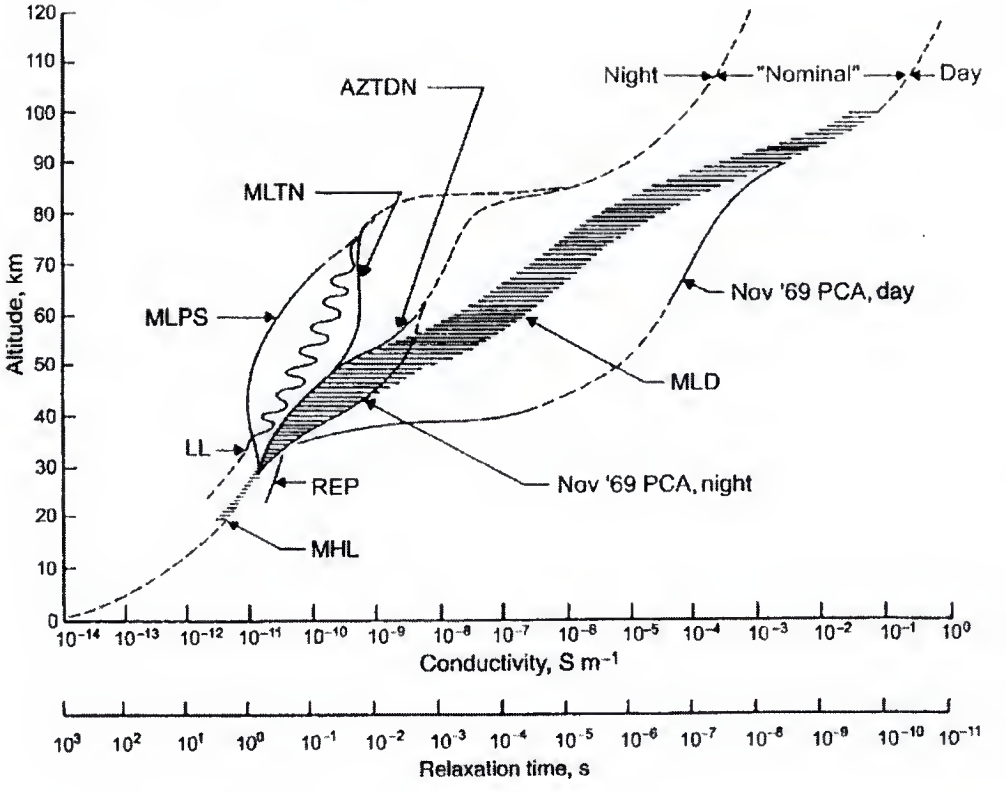

Figure 3–3: Electrical conductivity σ and corresponding relaxation time $\tau = \varepsilon\sigma^{-1}$, where $\varepsilon = 8.85^{-12}$ F m^{-1}, versus altitude under a variety of geophysical conditions. LL, low latitude, "wavy"; MLPS, mid-latitude typical night (high-latitude, quite); AZTDN, auroral zone, typical disturbed night; MLD, mid-latitude day, quite; MHL, mid-high-latitude, typical of $\sim$ 100 measurements; REP, relativistic electron (energy from a few MeV to 10 MeV) precipitation event (unusual); PCA, polar cap absorption event (an unusually large flux of energetic, $\sim$ 100 MeV) solar protons within the polar cap). Adapted from Hale [8]. (Source: reprinted with permission from Rakov and Uman [21], published in 2003, copyright Vladimir A Rakov and Martin A Uman, published by Cambridge University Press)

Lightning: physics and effects [21]. In the book [21], Rakov and Uman give a figure adapted from Hale [8] that estimates conductivity at altitude up to 120 km. Figure 1.3 in [21] is reproduced in Figure 3–3 of this thesis, with permission of the authors. In Appendix A.2, we give the following approximation based on Figure 3–3:

$$\sigma = 10^{p(z)},$$

where $p(z)$ is a quartic polynomial

$$p(z) = \sum_{i=0}^{4} a_i z^i, \tag{3.21}$$

where z is in kilometers and a_i are coefficients of the polynomial $p(z)$ as shown in Table 3-1. Then we have

Table 3–1: Coefficients of the polynomial $p(z)$

i	a_i
0	-1.316×10^{1}
1	6.351×10^{-2}
2	8.682×10^{-4}
3	8.674×10^{-6}
4	5.781×10^{-8}

$$\frac{\sigma_z}{\sigma} = (0.06351 + 0.001736z - 0.00002602z^2 + 0.0000002312z^3)(\ln 10)$$

and

$$\frac{\partial}{\partial z}\left(\frac{\sigma_z}{\sigma}\right) = (0.001736 - 0.00005204z + 0.0000006936z^2)(\ln 10).$$

For the approximations estimated from Figure 3–3, we can show that $\frac{\sigma_z}{\sigma}$ and $\frac{\partial}{\partial z}\left(\frac{\sigma_z}{\sigma}\right)$ are bounded in the domain Ω_1 (defined in (3.5)) as follows:

$$\left|\frac{\sigma_z}{\sigma}\right| \leq 0.4793$$

and

$$\left|\frac{\partial}{\partial z}\left(\frac{\sigma_z}{\sigma}\right)\right| \leq 0.007987.$$

3.4 Discussion: $d = \gamma_z(1 - e^{-\gamma(\Delta t)})/\gamma$

In this section, we discuss the parameter d, where

$$d = (1 - e^{-\gamma(\Delta t)})\frac{\gamma}{\gamma_z}, \tag{3.22}$$

$\gamma = \sigma/\varepsilon$, and $\gamma_z = \sigma_z/\varepsilon$. The parameter d is a critical factor when we discuss max-norm of some important matrices in Chapter 4.

Lemma 3.4.1 *For d as defined in (3.22), we have*

$$d = O(\Delta t),$$

while $d \leq 0.4793$ in Ω_1 for any Δt.

Proof: For any Δt, we have

$$\begin{aligned} d &= (1 - e^{-\gamma(\Delta t)})\frac{\gamma_z}{\gamma} \\ &\leq \frac{\gamma_z}{\gamma} \\ &\leq \frac{\sigma_z}{\sigma} \\ &\leq 0.4793. \end{aligned}$$

We also have

$$\begin{aligned} d &= \frac{1 - e^{-\gamma(\Delta t)}}{\gamma}\frac{\sigma_z}{\varepsilon} \\ &= \left(1 - e^{-\gamma(\Delta t)}\right)\frac{\sigma_z}{\sigma} \\ &= \gamma e^{-\gamma\xi\Delta t}\frac{\sigma_z}{\sigma}\Delta t, \end{aligned}$$

where $0 \leq \xi \leq 1$. Since $e^{-\gamma\xi\Delta t} \leq 1$, we have

$$d \leq \frac{\sigma_z}{\varepsilon}\Delta t = O(\Delta t)$$

■

Lemma 3.4.2 *For d as defined in (3.22) and $\gamma \in \mathcal{C}^2$, we have*

$$|d_i - d_{i+1}| = O(\Delta t \cdot h).$$

Proof:

$$\begin{aligned}
d_i - d_{i+1} &= \left(1 - e^{-\gamma_i(\Delta t)}\right)\left(\frac{\sigma_z}{\sigma}\right)_i - \left(1 - e^{-\gamma_{i+1}(\Delta t)}\right)\left(\frac{\sigma_z}{\sigma}\right)_{i+1} \\
&= \left(1 - e^{-\gamma_i(\Delta t)}\right)\left(\frac{\sigma_z}{\sigma}\right)_i - \left(1 - e^{-\gamma_i(\Delta t)}\right)\left(\frac{\sigma_z}{\sigma}\right)_{i+1} \\
&\quad + \left(1 - e^{-\gamma_i(\Delta t)}\right)\left(\frac{\sigma_z}{\sigma}\right)_{i+1} - \left(1 - e^{-\gamma_{i+1}(\Delta t)}\right)\left(\frac{\sigma_z}{\sigma}\right)_{i+1} \\
&= \left(1 - e^{-\gamma_i(\Delta t)}\right)\left[\left(\frac{\sigma_z}{\sigma}\right)_i - \left(\frac{\sigma_z}{\sigma}\right)_{i+1}\right] + \left(e^{-\gamma_{i+1}(\Delta t)} - e^{-\gamma_i(\Delta t)}\right)\left(\frac{\sigma_z}{\sigma}\right)_{i+1}
\end{aligned}$$

By the mean value theorem, we have

$$d_i - d_{i+1} = \left(1 - e^{-\gamma_i(\Delta t)}\right)\overline{\frac{\partial}{\partial z}\left(\frac{\sigma_z}{\sigma}\right)}_i h + e^{-\bar{\bar{\gamma}}_i(\Delta t)}\left(-\bar{\bar{\gamma}}_i'(\Delta t)\right) h \left(\frac{\sigma_z}{\sigma}\right)_{i+1},$$

where $\overline{\frac{\partial}{\partial z}\left(\frac{\sigma_z}{\sigma}\right)_i}$ is the derivative of $\left(\frac{\sigma_z}{\sigma}\right)$ with respect to z at some $\overline{z_i}$ between z_i and z_{i+1}, $\bar{\bar{\gamma}}_i$ and $\bar{\bar{\gamma}}_i'$ are the value and the derivative of γ at some $\overline{\overline{z_i}}$ between z_i and z_{i+1}. Since Ω is bounded and $\sigma \in \mathcal{C}^2$, both $\frac{\sigma_z}{\sigma}$ and $\frac{\partial}{\partial z}(\frac{\sigma_z}{\sigma})$ are bounded, and we have

$$\begin{aligned}
|d_i - d_{i+1}| &\leq \left|\left(1 - e^{-\gamma_i(\Delta t)}\right)\overline{\frac{\partial}{\partial z}\left(\frac{\sigma_z}{\sigma}\right)}_i h\right| + \left|\bar{\bar{\gamma}}_i'(\Delta t)\left(\frac{\sigma_z}{\sigma}\right)_{i+1} h\right| \\
&\leq \left|\gamma_i(\Delta t)\overline{\frac{\partial}{\partial z}\left(\frac{\sigma_z}{\sigma}\right)}_i h\right| + \left|\bar{\bar{\gamma}}_i'(\Delta t)\left(\frac{\sigma_z}{\sigma}\right)_{i+1} h\right| \\
&= O(\Delta t \cdot h).
\end{aligned}$$

■

CHAPTER 4
STABILITY AND ERROR ANALYSIS

4.1 Stability

This section uses Fourier analysis in our study of stability. We use Fourier analysis on the grid of integers, $\mathbf{Z}$ or $h\mathbf{Z}$, where $h\mathbf{Z}$ is defined by $h\mathbf{Z} = \{hm : m \in \mathbf{Z}\}$. Strikwerda [23] gives the Fourier transform as follows:

Definition 4.1.1 *If v is a grid function in ℓ^2 defined for all integers m, its Fourier transform is given by*

$$\hat{v}(\xi) = \frac{1}{\sqrt{2\pi}} \sum_{m=-\infty}^{\infty} e^{-im\xi} v_m$$

for $\xi \in [-\pi, \pi]$, and $\hat{v}(-\pi) = \hat{v}(\pi)$. If the spacing between the grid points is h, by changing variables we have

$$\hat{v}(\xi) = \frac{1}{\sqrt{2\pi}} \sum_{m=-\infty}^{\infty} e^{-imh\xi} v_m h$$

for $\xi \in [-\pi/h, \pi/h]$.

Next, we consider ψ a grid function whose spacing between the grid points is h. Then we have

1.

$$\begin{aligned}
& \frac{1}{\sqrt{2\pi}} \sum_{m=-\infty}^{\infty} e^{-imh\theta}(\psi_{m+1} - 2\psi_m + \psi_{m-1})h \\
= \; & \frac{1}{\sqrt{2\pi}} \sum_{m=-\infty}^{\infty} e^{-imh\theta}(e^{i\theta} - 2 + e^{-i\theta})\psi_m h \\
= \; & (2\cos\theta - 2)\frac{1}{\sqrt{2\pi}} \sum_{m=-\infty}^{\infty} e^{-imh\theta}\psi_m h \\
= \; & (2\cos\theta - 2)\hat{\psi}_m,
\end{aligned}$$

2.

$$\begin{aligned}
&\frac{1}{\sqrt{2\pi}}\sum_{m=-\infty}^{\infty} e^{-imh\theta}(\psi_{m+1}-\psi_{m-1})h \\
= &\frac{1}{\sqrt{2\pi}}\sum_{m=-\infty}^{\infty} e^{-imh\theta}(e^{i\theta}-e^{-i\theta})\psi_m h \\
= &\ 2i\sin\theta\frac{1}{\sqrt{2\pi}}\sum_{m=-\infty}^{\infty} e^{-imh\theta}\psi_m h \\
= &\ (2i\sin\theta)\hat{\psi}_m.
\end{aligned}$$

Theorem 4.1.2 *The finite difference scheme, Equation (3.12), is unconditionally stable when* $1/2 \le \mu \le 1$ *and is unstable when* $0 \le \mu < 1/2$.

Proof: Take Fourier transform of Equation (3.12) to obtain

$$\begin{aligned}
&\sum_{j=1}^{3}(2\cos\theta_j - 2)\hat{\psi}^{n+1} \\
= &\ \lambda_k\sum_{j=1}^{3}(2\cos\theta_j-2)\hat{\psi}^n - d_k h(1-\mu)i\sin\theta_3\hat{\psi}^n - d_k h\mu i\sin\theta_3\hat{\psi}^{n+1}. \\
\Longrightarrow\quad \hat{\psi}^{n+1} = &\ \frac{2\lambda_k\sum(\cos\theta_j-1) - d_k h(1-\mu)i\sin\theta_3}{2\sum(\cos\theta_j-1)+d_k h\mu i\sin\theta_3}\hat{\psi}^n.
\end{aligned}$$

Define the amplification factor as

$$\rho = \frac{2\lambda_k\sum(\cos\theta_j-1) - d_k h(1-\mu)i\sin\theta_3}{2\sum(\cos\theta_j-1)+d_k h\mu i\sin\theta_3}.$$

Then the magnitude of ρ is given by

$$|\rho|^2 = \frac{4\lambda_k^2\left(\sum(\cos\theta_j-1)\right)^2 + d_k^2h^2(1-\mu)^2\sin^2\theta_3}{4\left(\sum(\cos\theta_j-1)\right)^2 + d_k^2h^2\mu^2\sin^2\theta_3}.$$

For $1/2 \le \mu \le 1$, we have $|\rho|^2 \le 1$, since $\lambda_k = e^{-\gamma_k(\Delta t)} \le 1$. Hence the scheme is unconditionally stable. For $0 \le \mu < 1/2$, we want to show $|\rho|^2 \le 1 + K(\Delta t)$, where K is a constant, does not hold for any $\theta_j \in [-\pi, \pi]$. If $|\rho|^2 \le 1 + K(\Delta t)$, then we have

$$\frac{4\lambda_k^2\left(\sum(\cos\theta_j-1)\right)^2 + d_k^2h^2(1-\mu)^2\sin^2\theta_3}{4\left(\sum(\cos\theta_j-1)\right)^2 + d_k^2h^2\mu^2\sin^2\theta_3} \le 1 + K(\Delta t).$$

$$
\begin{aligned}
\iff \quad & 4\lambda_k^2 \left(\sum(\cos\theta_j - 1)\right)^2 + d_k^2 h^2 (1-\mu)^2 \sin^2\theta_3 \\
\leq \quad & 4\left(\sum(\cos\theta_j - 1)\right)^2 + 4K(\Delta t)\left(\sum(\cos\theta_j - 1)\right)^2 \\
& + d_k^2 h^2 \mu^2 \sin^2\theta_3 + K(\Delta t) d_k^2 h^2 \mu^2 \sin^2\theta_3 \\
\iff \quad & 4(\lambda_k^2 - 1 - K(\Delta t))\left(\sum(\cos\theta_j - 1)\right)^2 \leq d_k^2 h^2 (K(\Delta t)\mu^2 + 2\mu - 1)\sin^2\theta_3 \\
\iff \quad & 4(\lambda_k^2 - 1 - K(\Delta t)) \leq d_k^2 h^2 (K(\Delta t)\mu^2 + 2\mu - 1)\frac{\sin^2\theta_3}{\left(\sum(\cos\theta_j - 1)\right)^2} \\
\iff \quad & 4(1 + K(\Delta t)) - \lambda_k^2) \geq d_k^2 h^2 (1 - 2\mu - K(\Delta t)\mu^2)\frac{\sin^2\theta_3}{\left(\sum(\cos\theta_j - 1)\right)^2}
\end{aligned}
$$

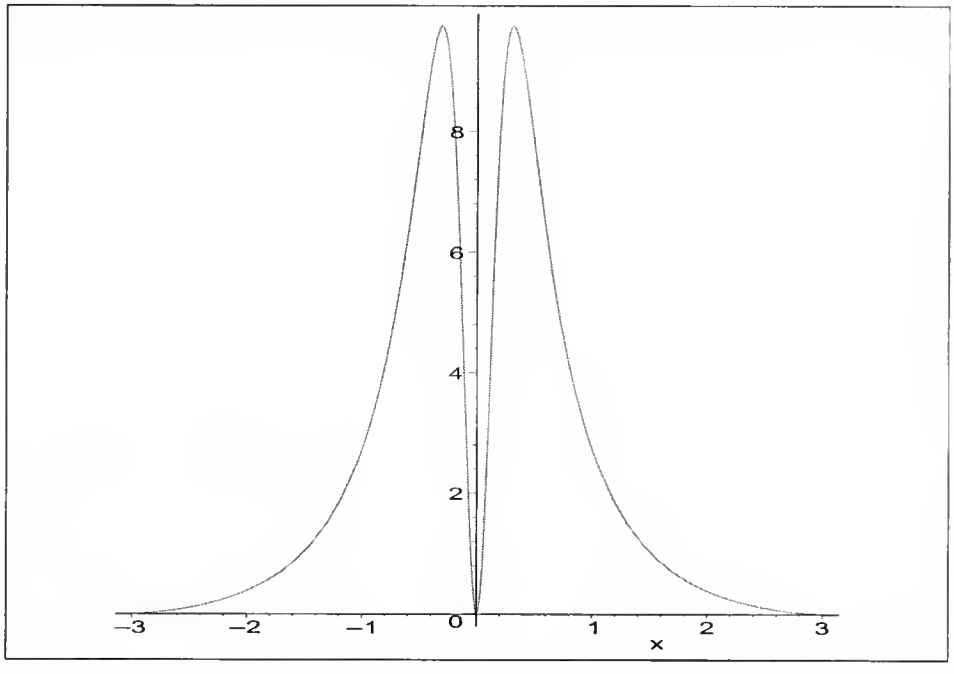

Figure 4–1: Graph of $f(x) = \dfrac{\sin^2 x}{(a+\cos x)^2}$, when $a = -1.05$.

Now consider the function

$$f(\theta_3) = \frac{\sin^2\theta_3}{\left(\sum(\cos\theta_j - 1)\right)^2} = \frac{\sin^2\theta_3}{(a+\cos\theta_3)^2},$$

where $a = \cos\theta_1 + \cos\theta_2 - 3$, and $-5 \leq a \leq -1$. Let

$$f(x) = \frac{\sin^2 x}{(a+\cos x)^2},$$

where $-\pi \leq x \leq \pi$, and $-5 \leq a \leq -1$. If $a \neq -1$, the graph of $f(x)$ is shown in Figure 4–1. We can find the maximum of $f(x)$ is $-a/(a^2-1)^{3/2}$. When a tends to -1, the maximum of $f(x)$ approaches to infinity. If $a = -1$, the graph of $f(x)$ is shown in Figure 4–2. We can see the supremum of $f(x)$ is infinity. Hence,

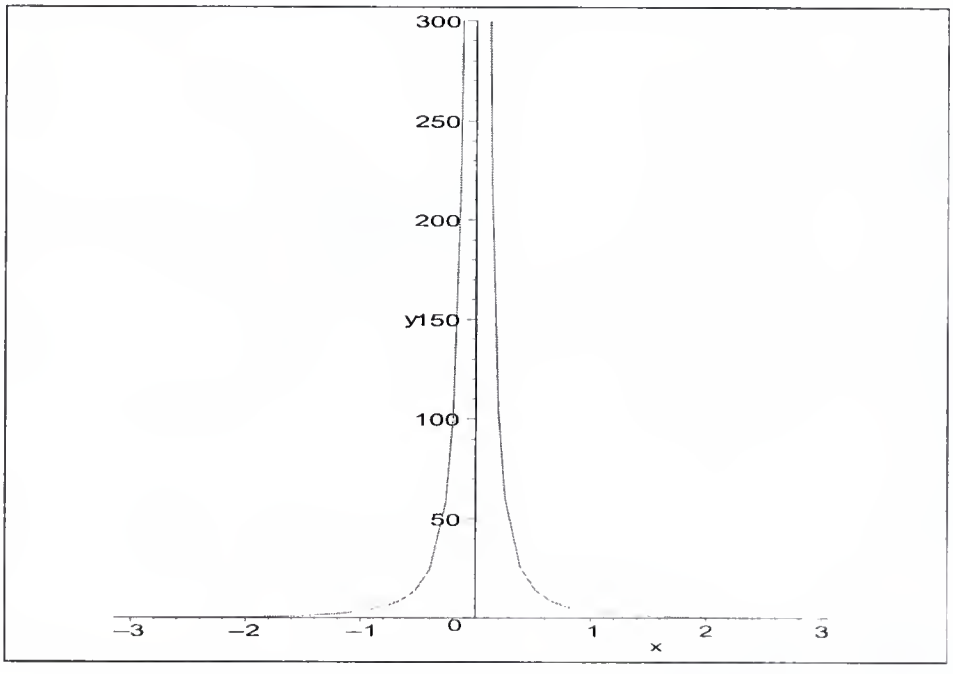

Figure 4–2: Graph of $f(x) = \dfrac{\sin^2 x}{(a+\cos x)^2}$, when $a = -1$.

$$\sup_{-\pi/h \leq \theta_3 \leq \pi/h} \frac{\sin^2 \theta_3}{(\sum(\cos\theta_j - 1))^2} = +\infty. \tag{4.1}$$

Therefore, it is impossible to have the inequality

$$4(1 + K(\Delta t)) - \lambda_k^2) \geq d_k^2 h^2 (1 - 2\mu - K(\Delta t)\mu^2) \frac{\sin^2 \theta_3}{(\sum(\cos\theta_j - 1))^2}.$$

In other words, the scheme is unstable when $0 \leq \mu < 1/2$. ■

4.2 Error Analysis

In this section, the error generated from the discretizing process is analyzed. Lemma 3.4.1 and Lemma 4.2.1 are applied in the error analysis. Lemma 3.4.1 gives

the estimate for d, and Hager [6] mentions the error estimate for integration as described in Lemma 4.2.1.

Lemma 4.2.1 *Let $W^{k,\infty}([a,b])$ denote the space of functions defined on the interval $[a,b]$ with derivatives through order k essentially bounded. If $g \in W^{1,\infty}$ and $f \in W^{2,\infty}$, then we have the estimate*

$$\left| \int_a^b f(z)g(z)dz - f(c)\int_a^b g(z)dz \right| \leq \frac{(b-a)^3}{12}\|f'\|\|g'\| + \frac{(b-a)^3}{24}\|f''\|\|g\|,$$

where $\|\cdot\|$ denotes the essential supremum and c is the midpoint of the interval $[a,b]$.

Proof:

$$\begin{aligned}
& \left| \int_a^b f(z)g(z)dz - f(c)\int_a^b g(z)dz \right| \\
= & \left| \int_a^b (f(z)-f(c))\, g(z)dx \right| \\
= & \left| \int_a^b \left(\int_c^z f'(s)ds \right) g(z)dz \right| \\
= & \left| \int_a^b \left(\int_c^z \left[f'(c) + \int_c^s f''(\tau)d\tau \right] ds \right) g(z)dz \right| \\
\leq & \left| \int_a^b \left(\int_c^z f'(c)ds \right) g(z)dz \right| + \left| \int_a^b \left(\int_c^z \int_c^s f''(\tau)d\tau ds \right) g(z)dz \right| \\
\leq & \; I_1 + I_2,
\end{aligned}$$

where

$$\begin{aligned}
I_1 &= \left| \int_a^b \left(\int_c^z f'(c)ds \right) \left(g(c) + \int_c^z g'(s)ds \right) dz \right| \\
&\leq \left| \int_a^b (z-c)f'(c)g(c)dz \right| + \left| \int_a^b (z-c)f'(c) \left(\int_c^z g'(s)ds \right) dz \right| \\
&\leq \|f'\|\|g'\| \left| \int_a^b (z-c)^2 dz \right| \\
&= \frac{(b-a)^3}{12}\|f'\|\|g'\|,
\end{aligned}$$

and

$$\begin{aligned} I_2 &\leq \|f''\|\|g\| \left| \int_a^b \left(\int_c^z \int_c^s d\tau ds \right) dz \right| \\ &= \frac{(b-a)^3}{24} \|f''\|\|g\|. \end{aligned}$$

This proves the lemma. ■

Note that γ is considered small in Lemma 4.2.2, Lemma 4.2.3, and Theorem 4.2.4.

Lemma 4.2.2 *If we use the finite difference scheme (3.12) to discretize the integral equation (3.10), then the truncation error is*

$$O(\Delta t)^m + O(h^2),$$

where $m = 3$ *if* $\mu = 1/2$ *and* $m = 2$ *if* $\mu \neq 1/2$,

Proof: Define R_1 as follows:

$$\begin{aligned} R_1 &= \int_{t_n}^{t_{n+1}} e^{\gamma(s-t_{n+1})} \gamma_z \frac{\partial \psi}{\partial z} ds - \gamma_z \left(\frac{\partial \psi}{\partial z} \right)^{n+\frac{1}{2}} \int_{t_n}^{t_{n+1}} e^{\gamma(s-t_{n+1})} ds \\ &= \int_{t_n}^{t_{n+1}} e^{\gamma(s-t_{n+1})} \gamma_z \frac{\partial \psi}{\partial z} ds - \frac{1 - e^{-\gamma(\Delta t)}}{\gamma} \gamma_z \left(\frac{\partial \psi}{\partial z} \right)^{n+\frac{1}{2}} \\ &= \int_{t_n}^{t_{n+1}} e^{\gamma(s-t_{n+1})} \gamma_z \frac{\partial \psi}{\partial z} ds - d \left(\frac{\partial \psi}{\partial z} \right)^{n+\frac{1}{2}}, \end{aligned}$$

where $d = \gamma_z(1 - e^{-\gamma(\Delta t)})/\gamma$. Apply Lemma 4.2.1 to obtain the following estimate:

$$\begin{aligned} |R_1| &= \left| \int_{t_n}^{t_{n+1}} e^{\gamma(s-t_{n+1})} \gamma_z \frac{\partial \psi}{\partial z} ds - \gamma_z \left(\frac{\partial \psi}{\partial z} \right)^{n+\frac{1}{2}} \int_{t_n}^{t_{n+1}} e^{\gamma(s-t_{n+1})} ds \right| \\ &\leq C(\Delta t)^3. \end{aligned} \tag{4.2}$$

Define R_2 as follows:

$$R_2 = d \left(\frac{\partial \psi}{\partial z} \right)^{n+\frac{1}{2}} - d \left[(1-\mu) \left(\frac{\partial \psi}{\partial z} \right)^n + \mu \left(\frac{\partial \psi}{\partial z} \right)^{n+1} \right].$$

where $0 \leq \mu \leq 1$. By Taylor's expansion, we have

$$\frac{\partial \psi}{\partial z}(t_n) = \frac{\partial \psi}{\partial z}(t_{n+\frac{1}{2}}) + \left(-\frac{\Delta t}{2}\right)\left(\frac{\partial \psi}{\partial z}\right)_t (t_{n+\frac{1}{2}}) + \frac{1}{2}\left(-\frac{\Delta t}{2}\right)^2 \left(\frac{\partial \psi}{\partial z}\right)_{tt} (t_{n+\frac{1}{2}}) + \cdots,$$

and

$$\frac{\partial \psi}{\partial z}(t_{n+1}) = \frac{\partial \psi}{\partial z}(t_{n+\frac{1}{2}}) + \left(\frac{\Delta t}{2}\right)\left(\frac{\partial \psi}{\partial z}\right)_t (t_{n+\frac{1}{2}}) + \frac{1}{2}\left(\frac{\Delta t}{2}\right)^2 \left(\frac{\partial \psi}{\partial z}\right)_{tt} (t_{n+\frac{1}{2}}) + \cdots.$$

Hence

$$\begin{aligned}(1-\mu)\frac{\partial \psi}{\partial z}(t_n) + \mu \frac{\partial \psi}{\partial z}(t_{n+1}) = & \\ \frac{\partial \psi}{\partial z}(t_{n+\frac{1}{2}}) + \left[(1-\mu)\left(-\frac{\Delta t}{2}\right) + \mu\left(\frac{\Delta t}{2}\right)\right]\left(\frac{\partial \psi}{\partial z}\right)_t (t_{n+\frac{1}{2}}) & \\ + \frac{1}{2}\left[(1-\mu)\left(-\frac{\Delta t}{2}\right)^2 + \mu\left(\frac{\Delta t}{2}\right)^2\right]\left(\frac{\partial \psi}{\partial z}\right)_{tt} (t_{n+\frac{1}{2}}) + \cdots &\end{aligned}$$

This implies that if $\mu = 1/2$,

$$(1-\mu)\left(\frac{\partial \psi}{\partial z}\right)^n + \mu\left(\frac{\partial \psi}{\partial z}\right)^{n+1} - \left(\frac{\partial \psi}{\partial z}\right)^{n+\frac{1}{2}} = O(\Delta t)^2,$$

and if $\mu \neq 1/2$,

$$(1-\mu)\left(\frac{\partial \psi}{\partial z}\right)^n + \mu\left(\frac{\partial \psi}{\partial z}\right)^{n+1} - \left(\frac{\partial \psi}{\partial z}\right)^{n+\frac{1}{2}} = O(\Delta t).$$

Therefore

$$|R_2| = O(\Delta t)^m, \tag{4.3}$$

where $m = 3$ if $\mu = 1/2$, and $m = 2$ if $\mu \neq 1/2$.

Let R_3 be defined as follows:

$$R_3 = \triangle \psi^{n+1} - \lambda \triangle \psi^n - \left(\triangle^h \psi^{n+1}_{i,j,k} - \lambda_k \triangle^h \psi^n_{i,j,k}\right).$$

By Taylor's expansion, we have

$$|R_3| = O(h^2). \tag{4.4}$$

Combining (4.2), (4.3), and (4.4), we see that when the solution to (3.10) is substituted in (3.12), we obtain

$$R_1 + R_2 + R_3 = O(\Delta t)^m + O(h^2),$$

where $m = 3$ if $\mu = 1/2$ and $m = 2$ if $\mu \neq 1/2$. ∎

Now if we add the force term $\mathbf{J}$ back to Equation (3.10), then the corresponding finite difference scheme is

$$\begin{aligned} \triangle^h \psi_{i,j,k}^{n+1} &= \lambda_k \triangle^h \psi_{i,j,k}^n - \frac{h}{2}\left[(1-\mu)\frac{d_k}{h^2}\partial_z \psi_{i,j,k}^n + \mu \frac{d_k}{h^2}\partial_z \psi_{i,j,k}^{n+1}\right] \\ &\quad - \frac{1 - e^{-\gamma_k(\Delta t)}}{\gamma_k}\frac{(\nabla \cdot \mathbf{J})^{n+\frac{1}{2}}}{\varepsilon}. \end{aligned} \tag{4.5}$$

Lemma 4.2.3 *If we use the finite difference scheme (4.5) to discretize Equation (3.7), then it has the truncation error*

$$O(\Delta t)^m + O(h^2),$$

where $m = 3$ if $\mu = 1/2$ and $m = 2$ if $\mu \neq 1/2$.

Proof: We do not include the force term $\mathbf{J}$ in Lemma 4.2.2. Now we only need to focus on the error caused by estimating the $\mathbf{J}$ term when the finite difference scheme (4.5) is applied to Equation (3.7). In the scheme (4.5), we approximate the $\nabla \cdot \mathbf{J}$ term in (3.7) by the $\nabla \cdot \mathbf{J}$ term in (4.5). By Lemma 4.2.1, we have

$$\left| \int_{t_n}^{t_{n+1}} e^{\gamma(s-t_{n+1})}\frac{\nabla \cdot \mathbf{J}}{\varepsilon} ds - \frac{(\nabla \cdot \mathbf{J})^{n+\frac{1}{2}}}{\varepsilon}\int_{t_n}^{t_{n+1}} e^{\gamma(s-t_{n+1})} ds \right| = O(\Delta t)^3.$$

Combining this estimate with Lemma 4.2.2, we find that when the solution to (3.7) is substituted in (4.5), we obtain

$$O(\Delta t)^m + O(h^2),$$

where $m = 3$ if $\mu = 1/2$, and $m = 2$ if $\mu \neq 1/2$. ■ The finite difference scheme (4.5) yields a matrix-vector relation of the form

$$\mathbf{A}\Psi^{n+1} - \mathbf{\Lambda}\mathbf{A}\Psi^n = \frac{h}{2}\left[(1-\mu)\mathbf{B}\Psi^n + \mu\mathbf{B}\Psi^{n+1}\right] - \frac{1-e^{-\gamma(\Delta t)}}{\gamma}\frac{1}{\varepsilon}\mathbf{I}^n, \tag{4.6}$$

where $\mathbf{I}^n$ corresponds to the term $\nabla \cdot \mathbf{J}$ of (4.5) evaluated at time level $n + \frac{1}{2}$.

The next theorem gives us the max-norm error estimate when the solution to (3.7) is substituted into Equation (4.6). Define the max-norm for a vector $\mathbf{x}$ as

$$\|\mathbf{x}\|_\infty = \max_{1\leq i\leq n} |x_i|,$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$, and define the max-norm for a matrix $\mathbf{M}$ as

$$\|\mathbf{M}\|_\infty = \max_{1\leq i\leq n} \sum_{j=1}^{n} |m_{i,j}|,$$

where $\mathbf{M} = (m_{i,j})$ is an $n \times n$ matrix.

Theorem 4.2.4 *Let Θ^n denote the continuous solution of (3.7) evaluated at $t = n\Delta t$ and at the mesh points of the grid $\{(x_i, y_j, z_k)\}$. Let Ψ^n be the vector with components $\psi_{i,j,k}$ in Equation (4.6). Then we have*

$$\|\Psi^n - \Theta^n\|_\infty = O(\Delta t)^m + O(h^2),$$

where $m = 2$ if $\mu = 1/2$ and $m = 1$ if $\mu \neq 1/2$.

Proof: Insert Θ^n into Equation (4.6) and we get

$$\mathbf{A}\Theta^{n+1} - \mathbf{\Lambda}\mathbf{A}\Theta^n = \frac{h}{2}\left[(1-\mu)\mathbf{B}\Theta^n + \mu\mathbf{B}\Theta^{n+1}\right] - \frac{1-e^{-\gamma(\Delta t)}}{\gamma}\frac{1}{\varepsilon}\mathbf{I}^n + \tau_n, \tag{4.7}$$

where τ_n is the truncation error and

$$\|\tau_n\|_\infty \leq O(\Delta t)^m + O(h^2),$$

where $m = 3$ if $\mu = 1/2$, and $m = 2$ if $\mu \neq 1/2$. Then by subtracting Equation (4.7) from Equation (4.6), we have

$$\boldsymbol{\Psi}^{n+1} - \boldsymbol{\Theta}^{n+1} = \mathbf{C}^{-1}\mathbf{D}\left(\boldsymbol{\Psi}^n - \boldsymbol{\Theta}^n\right) - \mathbf{C}^{-1}\tau_n, \tag{4.8}$$

where $\mathbf{C} = \mathbf{A} - \frac{h}{2}\mu\mathbf{B}$ and $\mathbf{D} = \boldsymbol{\Lambda}\mathbf{A} + \frac{h}{2}(1-\mu)\mathbf{B}$. Following from Equation (4.8), it is found

$$\boldsymbol{\Psi}^n - \boldsymbol{\Theta}^n = \mathbf{M}^n\left(\boldsymbol{\Psi}^0 - \boldsymbol{\Theta}^0\right) - \sum_{l=0}^{n-1}\mathbf{M}^l\mathbf{C}^{-1}\tau_{n-l-1}, \tag{4.9}$$

where $\mathbf{M} = \mathbf{C}^{-1}\mathbf{D}$. Note that we prove $\|\mathbf{M}\|_\infty = 1 + O(\Delta t)$ in Section 5.3 for $0 \leq \mu \leq 1$ if γ is small. Therefore, $\|\mathbf{M}\|_\infty$ can be written as $\|\mathbf{M}\|_\infty \leq 1 + c\Delta t \leq e^{c\Delta t}$. Taking the max-norm of Equation (4.9), then we have

$$\|\boldsymbol{\Psi}^n - \boldsymbol{\Theta}^n\|_\infty \leq e^{nc\Delta t}\left(\|\boldsymbol{\Psi}^0 - \boldsymbol{\Theta}^0\|_\infty + \|\mathbf{C}^{-1}\|_\infty \sum_{l=0}^{n-1}\|\tau_{n-l-1}\|_\infty\right).$$

Note that $\|\mathbf{C}^{-1}\|_\infty$ is bounded by a constant which is independent of h (see Lemma 5.1.2). Consequently, if $\|\boldsymbol{\Psi}^0 - \boldsymbol{\Theta}^0\|_\infty = O(h^2)$, then

$$\|\boldsymbol{\Psi}^n - \boldsymbol{\Theta}^n\|_\infty = O(\Delta t)^m + O(h^2),$$

where $m = 2$ if $\mu = 1/2$, and $m = 1$ if $\mu \neq 1/2$. ■

CHAPTER 5
CONVERGENCE ANALYSIS

5.1 Max-Norm Bound for Matrix $\mathbf{C}^{-1} = \left(\mathbf{A} - \frac{h}{2}\mu\mathbf{B}\right)^{-1}$

In this section , we obtain a max-norm bound for $\mathbf{C}^{-1} = \left(\mathbf{A} - \frac{h}{2}\mu\mathbf{B}\right)^{-1}$. where $\mathbf{A}$ and $\mathbf{B}$ are defined in Equation (3.14) and Equation (3.15), respectively. Lemma 5.1.1 from Varga's *Matrix Iterative Analysis* [28] is used to show that the matrix $\mathbf{C}$ is invertible and the max-norm of the matrix $\mathbf{C}^{-1}$ is bounded by a constant which is independent of h.

Lemma 5.1.1 *If $\mathbf{M} = (m_{i,j})$ is a real irreducibly diagonally dominant $n \times n$ matrix with $m_{i,j} \leq 0$ for $i \neq j$, and $m_{i,i} > 0$ for all $1 \leq i \leq n$, then $\mathbf{M}^{-1} > 0$, i.e., all elements of $\mathbf{M}^{-1}$ are positive.*

Lemma 5.1.2 *If h is sufficiently small, then the matrix $\mathbf{C}$ is invertible. Moreover, $\|\mathbf{C}^{-1}\|_\infty$ is bounded by a constant which is independent of h.*

Proof: The proof of this result is the same proof used in [6] for a slightly different matrix. Note that all the elements in the matrix $\mathbf{C}$ satisfy the conditions in Lemma 5.1.1 except the elements on the subdiagonal of $\mathbf{C}$. The elements on the subdiagonal have the form $-1 + \frac{h}{2}\mu d$, where d is a bounded positive number from Section 3.4, and $0 \leq \mu \leq 1$. In order to use Lemma 5.1.1, we must satisfy the condition

$$-1 + \frac{h}{2}\mu d \leq 0. \tag{5.1}$$

Accordingly, if h is sufficiently small, $\mathbf{C}^{-1} \geq \mathbf{0}$, i.e., the elements of $\mathbf{C}^{-1}$ are positive by Lemma 5.1.1. Hence, $\|\mathbf{C}^{-1}\|_\infty = \|\mathbf{C}^{-1}\mathbf{1}\|_\infty$, where $\mathbf{1}$ is the vector with every component equals to 1. Furthermore, $\|\mathbf{C}^{-1}\|_\infty \leq \|\mathbf{C}^{-1}\mathbf{F}\|_\infty$ if $\mathbf{F} \geq \mathbf{1}$.

Note that the matrix $\mathbf{C}$ is the coefficient matrix corresponding to the expression

$$-\triangle^h \psi_{i,j,k} - d_k \mu \partial_z \psi_{i,j,k}, \tag{5.2}$$

where $\triangle^h$ is the discrete Laplacian operator, ∂_z is the discrete $\partial\psi/\partial z$, and $d_k = \left(1 - e^{-\gamma_k(\Delta t)}\right) \cdot (\gamma_z/\gamma)_k$. Let

$$N^h = \{(i,j,k) : (x_i, y_j, z_k) \in \Omega\}.$$

Next, we show that $\|\mathbf{C}^{-1}\|_\infty \leq 2a^2$, where $a = hI$, and I is the maximum value of the index i associated with points in the set N^h. Let us define $\theta_{i,j,k} = -(hi)^2$. Observe that

$$\frac{1}{h^2}\left[h^2(i+1)^2 - 2h^2 i^2 + h^2(i-1)^2\right] = 2 \geq 1.$$

Moreover, define $\omega_{i,j,k} = (hi)^2$ for (i,j,k) outside N^h. If $(i,j,k) \in N^h$, we choose $\omega_{i,j,k}$ to satisfy

$$-\triangle^h \omega_{i,j,k} - d_k \mu \partial_z \omega_{i,j,k} = 0. \tag{5.3}$$

Then from Equation (5.3), we obtain

$$\begin{aligned} \omega_{i,j,k} &= \frac{1}{6}\left[\omega_{i+1,j,k} + \omega_{i-1,j,k} + \omega_{i,j+1,k} + \omega_{i,j-1,k} + \left(1 + \frac{h}{2} d_k \mu\right)\omega_{i,j,k+1}\right. \\ &\quad \left. + \left(1 - \frac{h}{2} d_k \mu\right)\omega_{i,j,k-1}\right] \end{aligned} \tag{5.4}$$

for $(i,j,k) \in N^h$. If h is sufficiently small, the condition (5.1) is satisfied, that is, $1 - \frac{h}{2} d_k \mu \geq 0$. Hence, by Equation (5.4), we have

$$\max_{N^h} |\omega_{i,j,k}| \leq \max_{\partial N^h} |\omega_{i,j,k}|,$$

where ∂N^h is the set of boundary points of N^h. Hence,

$$|\omega_{i,j,k}| \leq (hi)^2 \leq (hI)^2 = a^2,$$

for all $(i,j,k) \in N^h$. If $\psi_{i,j,k} = \theta_{i,j,k} + \omega_{i,j,k}$, then we have

$$\begin{aligned} \psi_{i,j,k} &= 0 \qquad \text{if } (i,j,k) \in \partial N^h, \\ -\triangle^h \psi_{i,j,k} - d_k \mu \partial_z \psi_{i,j,k} &\geq 1 \qquad \text{if } (i,j,k) \in N^h. \end{aligned}$$

Let $\mathbf{\Psi}$, $\mathbf{\Theta}$, and $\mathbf{\Omega}$ denote the vectors formed from $\psi_{i,j,k}$, $\theta_{i,j,k}$, and $\omega_{i,j,k}$ for $(i,j,k) \in N^h$. Take $\mathbf{F} = \mathbf{C\Psi} \geq \mathbf{1}$ to obtain

$$\begin{aligned} \|\mathbf{C}^{-1}\|_\infty &\leq \|\mathbf{C}^{-1}\mathbf{F}\|_\infty = \|\mathbf{\Psi}\|_\infty \\ &= \|\mathbf{\Theta} + \mathbf{\Omega}\|_\infty \\ &\leq \|\mathbf{\Theta}\|_\infty + \|\mathbf{\Omega}\|_\infty \\ &\leq a^2 + a^2 = 2a^2, \end{aligned}$$

where a is a constant which is independent of h. ■

Remark: As noted in the condition (5.1), the matrix $\mathbf{C}$ will satisfy the conditions of Lemma 5.1.1 if h is sufficiently small. In fact, using the estimate for d from Section 3.4, we obtain the condition of $h \leq 4.173$ kilometers when $\mu = 1$ for the domain Ω_1.

5.2 Estimates for $\mathbf{A_1}^{-1}$

In this section, we will get some preliminary results for finding the max-norm of the amplification matrix $\mathbf{C}^{-1}\mathbf{D}$.

Lemma 5.2.1 *Let $\mathbf{A_1}$ be the $n \times n$ symmetric tridiagonal matrix with 2's on the diagonal and with -1's on the superdiagonal and on the subdiagonal. Then the i^{th} row of the inverse matrix of $\mathbf{A_1}$ has the form*

$$(\mathbf{A_1}^{-1})_i = \frac{1}{n+1}[n-i+1, 2(n-i+1), \cdots, i(n-i+1), i(n-i), i(n-i-1), \cdots, 2i, i].$$

Proof: Note that we compute the $(i,j)^{th}$ entry of the matrix $\mathbf{A_1}^{-1}\mathbf{A_1}$ by the formula

$$\begin{aligned}(\mathbf{A_1}^{-1}\mathbf{A_1})_{i,j} &= \sum_{k=1}^{n}(\mathbf{A_1}^{-1})_{i,k}(\mathbf{A_1})_{k,j} \\ &= (a^{-1})_{i,j-1}a_{j-1,j} + (a^{-1})_{i,j}a_{j,j} + (a^{-1})_{i,j+1}a_{j+1,j},\end{aligned}$$

where $(a^{-1})_{i,j}$ denotes the $(i,j)^{th}$ entry of the matrix $\mathbf{A_1}^{-1}$. For $j = i$, we have

$$(\mathbf{A_1}^{-1}\mathbf{A_1})_{i,i} = \frac{1}{n+1}[(i-1)(n-i+1)(-1) + i(n-i+1)(2) + i(n-i)(-1)] = 1.$$

For $j \neq i$, we consider two cases:

$$(\mathbf{A_1}^{-1}\mathbf{A_1})_{i,j} = \frac{1}{n+1}[(j-1)(n-i+1)(-1)+j(n-i+1)(2)+(j+1)(n-i+1)(-1)] = 0,$$

if $j = 1, 2, \cdots, i-1$, and

$$(\mathbf{A_1}^{-1}\mathbf{A_1})_{i,j} = \frac{1}{n+1}[i(n-i-(m-1))(-1)+i(n-i-m)(2)+i(n-i-(m+1))(-1)] = 0,$$

if $j = i+m$, where $m = 1, 2, \cdots, n-i$. Therefore, $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$. ■

Remark: Dr. J. Gopalakrishnan suggests to use the finite element method to find $\mathbf{A_1}^{-1}$. Consider the two-point value problem:

$$\begin{cases} -U'' = f \text{ on } (0,1), \\ U(0) = U(1) = 0. \end{cases}$$

Then the Green's function $G(x,s)$ is given by

$$G(x,s) = \begin{cases} (1-s)x & , \text{ if } x < s, \\ (1-x)s & , \text{ if } x > s. \end{cases}$$

Let $0 = x_0 < x_1 < \cdots < x_{n+1} = 1$ with $x_{i+1} - x_i = 1/(n+1)$ for all $i = 0, 1, \ldots, n$. Define

$$a(u,v) = \int_0^1 u'(x)v'(x)dx,$$

and $S = \{v : v(x)$ is continuous on [0,1], linear on each interval $[x_i, x_{i+1}]$ for all $i = 0, 1, \ldots n$, and $v(0) = v(1) = 0\}$. The discrete Green's function $G_S^j(x) \in S$ is defined by

$$a(G_S^j, v) = v(x_j) \quad \text{for all } v \in S, \tag{5.5}$$

for any grid vertex x_j. Then

$$G_S^j(x) := G(x, x_j) = \begin{cases} (1 - x_j)x & , \text{ if } x \leq x_j, \\ (1 - x)x_j & , \text{ if } x \geq x_j. \end{cases},$$

and

$$\begin{aligned} G(x_i, x_j) &= \frac{i(n+1-j)}{(n+1)^2}, \text{ if } i \leq j \\ G(x_i, x_j) &= \frac{j(n+1-i)}{(n+1)^2}, \text{ if } i > j \end{aligned}$$

Let $\{\phi_i\}_{i=0}^{n+1}$ be a basis of S, where

$$\phi_i(x) = \begin{cases} 1 & \text{if } x = x_i, \\ 0 & \text{if } x = x_j \ (j \neq i) \end{cases}$$

Then the discrete Green function can be written as

$$\begin{aligned} G_S^j(x) &= (1 - x_j) \sum_{k=0}^{j} \phi_k x_k + \left(\sum_{k=j+1}^{n+1} \phi_k - \sum_{k=j+1}^{n+1} \phi_k x_k \right) x_j \\ &= (1 - x_j) \sum_{k=1}^{j} \phi_k x_k + \left(\sum_{k=j+1}^{n} \phi_k - \sum_{k=j+1}^{n} \phi_k x_k \right) x_j, \end{aligned}$$

since $x_0 = 0$ and $x_{n+1} = 1$. Thus

$$
\begin{aligned}
a(G_S^j, \phi_i) &= a\left((1-x_j)\sum_{k=1}^{j}\phi_k x_k + x_j\left(\sum_{k=j+1}^{n}\phi_k - \sum_{k=j+1}^{n}\phi_k x_k\right), \phi_i\right) \\
&= \sum_{k=1}^{j}[a(\phi_k,\phi_i)(1-x_j)x_k] + \sum_{k=j+1}^{n}[a(\phi_k,\phi_i)x_j(1-x_k)] \\
&= (a(\phi_i,\phi_1), a(\phi_i,\phi_2), \ldots, a(\phi_i,\phi_n))\begin{pmatrix} (1-x_j)x_1 \\ (1-x_j)x_2 \\ \vdots \\ (1-x_j)x_k \\ x_j(1-x_{k+1}) \\ x_j(1-x_{k+2}) \\ \vdots \\ x_j(1-x_n) \end{pmatrix} \qquad (5.6)
\end{aligned}
$$

By Equation (5.5), we know

$$a(G_S^j, \phi_i) = \phi_i(x_j) = \delta_{ij},$$

where $\delta_{i,j} = 1$ if $i = j$, and $\delta_{i,j} = 0$ otherwise. Note that

$$a(\phi_i, \phi_j) = \begin{cases} 2(n+1) & \text{if } j = i \\ -(n+1) & \text{if } j = i-1 \text{ or } j = i+1 \\ 0 & \text{otherwise} \end{cases}$$

Let $\mathbf{G}$ be the matrix with the $(i,j)^{th}$ entry $G(x_i, x_j)$. Equation (5.6) can be written as

$$(n+1)\mathbf{A_1}(i,\cdot)\cdot\mathbf{G}(\cdot,j) = \delta_{ij},$$

where $\mathbf{A_1}(i,\cdot)$ is the i^{th} row of the matrix $\mathbf{A_1}$ and $\mathbf{G}(\cdot,j)$ is the j^{th} column of the matrix $\mathbf{G}$. Then, we have

$$\mathbf{A_1}^{-1} = (n+1)\mathbf{G}.$$

Lemma 5.2.2 *Let* $\mathbf{A_1}$ *be the matrix described in Lemma 5.2.1. Let* $\mathbf{T}$ *be the* $n \times n$ *tridiagonal matrix with 0's on the diagonal, with 1's on the superdiagonal, and with -1's on the subdiagonal. Then* $\|\mathbf{A_1}^{-1}\mathbf{T}\|_\infty = n-1$.

Proof: Since

$$\begin{aligned}(\mathbf{A_1}^{-1}\mathbf{T})_{i,j} &= \sum_{k=1}^{n}(\mathbf{A_1}^{-1})_{i,k}(\mathbf{T})_{k,j} \\ &= (a^{-1})_{i,j-1}t_{j-1,j} + (a^{-1})_{i,j+1}t_{j+1,j} \\ &= (a^{-1})_{i,j-1} - (a^{-1})_{i,j+1} \quad ,\end{aligned}$$

we have

$$(\mathbf{A_1}^{-1}\mathbf{T})_{i,j} = \begin{cases} \dfrac{-2(n+1-i)}{n+1}, & \text{if } j = 1, 2, \cdots, i-1 \\ \dfrac{-n+2i-1}{n+1}, & \text{if } j = i \\ \dfrac{-2i}{n+1}, & \text{if } j = i+1, i+2, \cdots, n-1 \\ \dfrac{2i}{n+1}, & \text{if } j = n \end{cases}$$

Then, the infinity norm for the matrix $\mathbf{A_1}^{-1}\mathbf{T}$ is

$$\begin{aligned}\|\mathbf{A_1}^{-1}\mathbf{T}\|_\infty &= \max_i \sum_{j=1}^{n} \left|(\mathbf{A_1}^{-1}\mathbf{T})_{i,j}\right| \\ &= \max_i \left[\frac{2(n-i+1)}{n+1}(i-1) + \frac{|-n+2i-1|}{n+1} + \frac{2i}{n+1}(n-i)\right] \\ &= \max_i \left[\frac{-4i^2+4i(n+1)-2(n+1)}{n+1} + \frac{|2i-(n+1)|}{n+1}\right]. \qquad (5.7)\end{aligned}$$

Now focus on the term inside brackets of (5.7):

$$\begin{aligned}
& \frac{-4i^2+4i(n+1)-2(n+1)}{n+1}+\frac{|2i-(n+1)|}{n+1} \\
= & \frac{-[4i^2-4i(n+1)+(n+1)^2]+(n+1)^2-2(n+1)}{n+1}+\frac{|2i-(n+1)|}{n+1} \\
= & \frac{-[2i-(n+1)]^2+|2i-(n+1)|}{n+1}+(n-1) \qquad (5.8) \\
\leq & \ n-1,
\end{aligned}$$

for all $1 \leq i \leq n$, where i and n are both integers. The maximum value of Equation (5.8) is $n-1$ when

$$-[2i-(n+1)]^2+|2i-(n+1)|=0. \qquad (5.9)$$

Now we consider two cases:

Case 1: n is an odd number. The condition (5.9) is satisfied if we choose $i=(n+1)/2$. Putting this i into Equation (5.7), we obtain $n-1$.

Case 2: n is an even number. The condition (5.9) is satisfied if we choose $i=n/2$. Putting this i into Equation (5.7), we obtain $n-1$. ∎

Lemma 5.2.3 *Let* $f_i = e^{-\gamma_{i+1}(\Delta t)} - e^{-\gamma_i(\Delta t)}$, *and* $\gamma \in \mathcal{C}^2$. *Then* $|f_i| = O(\Delta t \cdot h)$ *and* $|f_i - f_{i+1}| = O(\Delta t \cdot h^2)$.

Proof: Both equalities are proved by the mean value theorem:

(1)

$$\begin{aligned}
f_i &= e^{-\gamma_{i+1}(\Delta t)} - e^{-\gamma_i(\Delta t)} \\
&= e^{-\overline{\gamma}_i(\Delta t)}(-\overline{\gamma}_i'(\Delta t))h,
\end{aligned}$$

where $\overline{\gamma}_i$ and $\overline{\gamma}_i'$ are the value and the derivative of γ at some $\overline{z}_i$ between z_i and z_{i+1}. Hence,

$$|f_i| = O(\Delta t \cdot h).$$

(2)

$$\begin{aligned} f_i - f_{i+1} &= e^{-\gamma_{i+1}(\Delta t)} - e^{-\gamma_i(\Delta t)} - \left(e^{-\gamma_{i+2}(\Delta t)} - e^{-\gamma_{i+1}(\Delta t)}\right) \\ &= e^{-\overline{\gamma}_i(\Delta t)}(-\overline{\gamma}_i'(\Delta t)h) - e^{-\overline{\gamma}_{i+1}(\Delta t)}(-\overline{\gamma}_{i+1}'(\Delta t)h) \\ &= (\Delta t)h\left(e^{-\overline{\gamma}_{i+1}(\Delta t)}\overline{\gamma}_{i+1}' - e^{-\overline{\gamma}_i(\Delta t)}\overline{\gamma}_i'\right) \\ &= (\Delta t)h\left(e^{-\overline{\overline{\gamma}}_i(\Delta t)}(-\overline{\overline{\gamma}}_i'(\Delta t))\overline{\overline{\gamma}}_i' h + e^{-\overline{\overline{\gamma}}_i(\Delta t)}\overline{\overline{\gamma}}_i'' h\right), \end{aligned}$$

where $\overline{\gamma}_i$ and $\overline{\gamma}_i'$ are the value and the derivative of γ at some $\overline{z}_i$ between z_i and z_{i+1}, $\overline{\gamma}_{i+1}$ and $\overline{\gamma}_{i+1}'$ are the value and the derivative of γ at some $\overline{z}_{i+1}$ between z_{i+1} and z_{i+2}, and $\overline{\overline{\gamma}}_i$, $\overline{\overline{\gamma}}_i'$, and $\overline{\overline{\gamma}}_i''$ are the value, the first derivative, and the second derivative of γ at some $\overline{\overline{z}}_i$ between z_i and z_{i+2}. Hence,

$$\begin{aligned} |f_i - f_{i+1}| &\leq (\Delta t)h\left((\overline{\overline{\gamma}}_i')^2(\Delta t)h + \overline{\overline{\gamma}}_i'' h\right) \\ &= O(\Delta t \cdot h^2). \end{aligned}$$

■

5.3 Max-Norm Bound for Matrix $\mathbf{C}^{-1}\mathbf{D}$

The max-norm bound for the matrix $\mathbf{C}^{-1}\mathbf{D}$, where $\mathbf{C} = \mathbf{A} - \frac{h}{2}\mu\mathbf{B}$, and $\mathbf{D} = \mathbf{\Lambda A} + \frac{h}{2}(1-\mu)\mathbf{B}$ are discussed in this section. Only the case $\mu = 1$ is discussed for the convergence analysis. Note that

$$\begin{aligned} \mathbf{C}^{-1}\mathbf{D} &= \left(\mathbf{A} - \frac{h}{2}\mathbf{B}\right)^{-1}(\mathbf{\Lambda A}) \\ &= \left(\mathbf{A} - \frac{h}{2}\mathbf{B}\right)^{-1}\mathbf{A\Lambda} + \left(\mathbf{A} - \frac{h}{2}\mathbf{B}\right)^{-1}(\mathbf{\Lambda A} - \mathbf{A\Lambda}) \\ &= \left(\mathbf{A} - \frac{h}{2}\mathbf{B}\right)^{-1}\left(\mathbf{A} - \frac{h}{2}\mathbf{B}\right)\mathbf{\Lambda} + \left(\mathbf{A} - \frac{h}{2}\mathbf{B}\right)^{-1}\frac{h}{2}\mathbf{B\Lambda} \\ &\quad + \left(\mathbf{A} - \frac{h}{2}\mathbf{B}\right)^{-1}(\mathbf{\Lambda A} - \mathbf{A\Lambda}) \\ &= \mathbf{\Lambda} + \frac{h}{2}\left(\mathbf{I} - \frac{h}{2}\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{A}^{-1}\mathbf{B\Lambda} + \left(\mathbf{I} - \frac{h}{2}\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{A}^{-1}(\mathbf{\Lambda A} - \mathbf{A\Lambda}) \end{aligned}$$

Hence,

$$\|\mathbf{C}^{-1}\mathbf{D}\|_\infty \leq \bar{\lambda} + \bar{\lambda}\frac{h}{2}\frac{\|\mathbf{A}^{-1}\mathbf{B}\|_\infty}{1-\frac{h}{2}\|\mathbf{A}^{-1}\mathbf{B}\|_\infty} + \frac{\|\mathbf{A}^{-1}(\mathbf{\Lambda A} - \mathbf{A\Lambda})\|_\infty}{1-\frac{h}{2}\|\mathbf{A}^{-1}\mathbf{B}\|_\infty}$$

whenever $\frac{h}{2}\mu\|\mathbf{A}^{-1}\mathbf{B}\|_\infty < 1$, and where $\bar{\lambda} = \max\lambda_i$. Our goal in this section is to show that $\|\mathbf{C}^{-1}\mathbf{D}\|_\infty = 1 + O(\Delta t)$.

5.3.1 Max-Norm for Matrix $\mathbf{C}^{-1}\mathbf{D}$ in 1-Dimensional Space

The preliminary results from Section 5.2 are used to show that $\|\mathbf{C}^{-1}\mathbf{D}\|_\infty = 1 + O(\Delta t)$ in one dimensional space.

Lemma 5.3.1 *Let* $\mathbf{A_1}$ *be the matrix described in Lemma 5.2.1. Let* $\mathbf{B_1}$ *be the* $n \times n$ *tridiagonal matrix with 0's on the diagonal, with* d_i*'s on the superdiagonal, and with* $-d_i$*'s on the subdiagonal. Then*

$$h\|\mathbf{A_1}^{-1}\mathbf{B_1}\|_\infty = O(\Delta t).$$

Proof: The matrix $\mathbf{B_1}$ has the form

$$\mathbf{B_1} = \begin{pmatrix} 0 & d_1 & & & & \\ -d_2 & 0 & d_2 & & & \\ & -d_3 & 0 & d_3 & & \\ & & -d_4 & 0 & d_4 & \\ & & & \cdot & \cdot & \cdot \\ & & & & \cdot & \cdot \end{pmatrix},$$

where each $d_i = O(\Delta t)$. We can write the matrix $\mathbf{B_1}$ as

$$\mathbf{B_1} = \mathbf{B}' + \mathbf{B}''$$

$$= \begin{pmatrix} 0 & d_1 & & & & & \\ 0 & 0 & d_2 & & & & \\ & -d_1 & 0 & d_3 & & & \\ & & -d_2 & 0 & d_4 & & \\ & & & \cdot & \cdot & \cdot & \\ & & & & \cdot & \cdot & \cdot \end{pmatrix} + \begin{pmatrix} 0 & 0 & & & & & \\ -d_2 & 0 & 0 & & & & \\ & d_1 - d_3 & 0 & 0 & & & \\ & & d_2 - d_4 & 0 & 0 & & \\ & & & & \cdot & \cdot & \cdot \\ & & & & & \cdot & \cdot \end{pmatrix}.$$

We will compute $\mathbf{A_1}^{-1}\mathbf{B}'$ and $\mathbf{A_1}^{-1}\mathbf{B}''$ respectively.

(1) compute $\mathbf{A_1}^{-1}\mathbf{B}'$:

$$(\mathbf{A_1}^{-1}\mathbf{B}')_{i,j} = d_{j-1}\left((a^{-1})_{i,j-1} - (a^{-1})_{i,j+1}\right),$$

where $(a^{-1})_{i,j}$ is the $(i,j)^{th}$ entry of $\mathbf{A_1}^{-1}$. Hence,

$$(\mathbf{A_1}^{-1}\mathbf{B}')_{i,j} = \begin{cases} 0 & \text{if } j = 1 \\ \dfrac{-2(n+i-1)}{n+1} d_{j-1} & \text{if } 2 \le j \le i-1 \\ \dfrac{-n+2i-1}{n+1} d_{j-1} & \text{if } j = i \\ \dfrac{2i}{n+1} d_{j-1} & \text{if } (i-1) \le j \le n \end{cases}$$

Therefore, for any i,

$$\begin{aligned} \sum_{j=1}^{n} |(\mathbf{A_1}^{-1}\mathbf{B}')_{i,j}| &= \frac{2(n-i+1)}{n+1}(d_1 + d_2 + \cdots + d_{i-2}) + \frac{|-n+2i-1|}{n+1} d_{i-1} \\ &\quad + \frac{2i}{n+1}(d_i + d_{i+1} + \cdots + d_{n-1}) \\ &\le 2(i-2)d + 3d + 2(n-i)d \text{ , where } d = \max d_i = O(\Delta t) \\ &= O\left(\frac{\Delta t}{h}\right). \end{aligned}$$

(2) compute $\mathbf{A_1}^{-1}\mathbf{B}''$:

$$(\mathbf{A_1}^{-1}\mathbf{B}'')_{i,j} = \begin{cases} -(a^{-1})_{i,2}d_2 & \text{if } j=1 \\ (a^{-1})_{i,j}(d_{j-1}-d_{j+1}) & \text{if } 1<j<n \\ 0 & \text{if } j=n \end{cases}$$

$$= \begin{cases} \dfrac{-2(n-i+1)}{n+1}d_2 & \text{if } j=1 \\ \dfrac{(j+1)(n-i+1)}{n+1}(d_{j-1}-d_{j+1}) & \text{if } 1<j<i \\ \dfrac{i(n-i-k)}{n+1}(d_{j-1}-d_{j+1}) & \text{if } j=i+k \ (k=0,1,\cdots,n-i-1) \\ 0 & \text{if } j=n \end{cases}$$

Hence, for any i,

$$\begin{aligned} \sum_{j=1}^{n}|(\mathbf{A_1}^{-1}\mathbf{B}'')_{i,j}| &\le \frac{2(n-i+1)}{n+1}d_2+\frac{n-i+1}{n+1}(3+4+\cdots+i)|d'| \\ &\quad +\frac{i}{n+1}\left[(n-i)+(n-i-1)+\cdots+2+1\right]|d'|, \\ &\quad \text{where } |d'|=\max|d_{j-1}-d_{j+1}|=O(\Delta t\cdot h) \\ &\le O(\Delta t)+\frac{(i+3)(i-2)}{2}O(\Delta t\cdot h)+\frac{(n-i+1)(n-i)}{2}O(\Delta t\cdot h) \\ &= O\left(\frac{\Delta t}{h}\right). \end{aligned}$$

Therefore,

$$\begin{aligned} h\|(\mathbf{A_1}^{-1}\mathbf{B_1})\|_\infty &\le h\left(\|(\mathbf{A_1}^{-1}\mathbf{B}')\|_\infty+\|(\mathbf{A_1}^{-1}\mathbf{B}'')\|_\infty\right) \\ &= O(\Delta t). \end{aligned}$$

■

Lemma 5.3.2 *Let* $\mathbf{A_1}$ *be the matrix described in Lemma 5.2.1 and* $\mathbf{\Lambda_1}$ *be the diagonal matrix with* λ_i *on the diagonal. Then*

$$\|\mathbf{A_1}^{-1}(\mathbf{\Lambda_1}\mathbf{A_1} - \mathbf{A_1}\mathbf{\Lambda_1})\|_\infty = O(\Delta t).$$

Proof: To prove this Lemma, we compute the max-norm of $\mathbf{A_1}^{-1}(\mathbf{\Lambda_1}\mathbf{A_1} - \mathbf{A_1}\mathbf{\Lambda_1})$. Let $\mathbf{F} = \mathbf{\Lambda_1}\mathbf{A_1} - \mathbf{A_1}\mathbf{\Lambda_1}$. Then we can write $\mathbf{F}$ as

$$\begin{aligned}
\mathbf{F} &= \begin{pmatrix} 0 & f_1 & & & \\ -f_1 & 0 & f_2 & & \\ & -f_2 & 0 & f_3 & \\ & & \cdot & \cdot & \cdot \\ & & & \cdot & \cdot \end{pmatrix}, \text{ where } f_i = e^{-\gamma_{i+1}(\Delta t)} - e^{-\gamma_i(\Delta t)} \\
&= \begin{pmatrix} 0 & f_2 & & & \\ -f_1 & 0 & f_3 & & \\ & -f_2 & 0 & f_4 & \\ & & \cdot & \cdot & \cdot \\ & & & \cdot & \cdot \end{pmatrix} + \begin{pmatrix} 0 & f_1 - f_2 & & & \\ 0 & 0 & f_2 - f_3 & & \\ & 0 & 0 & f_3 - f_4 & \\ & & \cdot & \cdot & \cdot \\ & & & \cdot & \cdot \end{pmatrix} \\
&= \mathbf{F}' + \mathbf{F}''.
\end{aligned}$$

The same process in Lemma 5.3.1 is used to compute $\mathbf{A_1}^{-1}\mathbf{F}'$ and $\mathbf{A_1}^{-1}\mathbf{F}''$ respectively.

(1) compute $\mathbf{A_1}^{-1}\mathbf{F}'$:

$$\begin{aligned}
(\mathbf{A_1}^{-1}\mathbf{F}')_{i,j} &= f_j\left((a^{-1})_{i,j-1} - (a^{-1})_{i,j+1}\right) \\
&= \begin{cases} \dfrac{-2(n-i+1)}{n+1} f_j & \text{if } j < i \\ \dfrac{(-n+2i+1)}{n+1} f_j & \text{if } j = i \\ \dfrac{2i}{n+1} f_j & \text{if } j > i \end{cases}
\end{aligned}$$

Therefore, for any i,

$$\begin{aligned}
\sum_{j=1}^{n} |(\mathbf{A_1}^{-1}\mathbf{F}')_{i,j}| &= \frac{2(n-i+1)}{n+1}(|f_1| + |f_2| + \cdots + |f_{i-1}|) + \frac{|-n+2i-1|}{n+1}|f_i| \\
&\quad + \frac{2i}{n+1}(|f_{i+1}| + |f_{i+2}| + \cdots + |f_n|) \\
&\leq 2(i-1)|f| + 3|f| + 2(n-i)|f|, \text{ where } |f| = \max|f_i| = O(\Delta t \cdot h) \\
&= (2n+1)O(\Delta t \cdot h) \\
&= O(\Delta t).
\end{aligned}$$

(2) compute $\mathbf{A_1}^{-1}\mathbf{F}''$:

$$(\mathbf{A_1}^{-1}\mathbf{F}'')_{i,j} = \begin{cases} 0 & \text{if } j = 1 \\ (a^{-1})_{i,j-1}(f_{j-1} - f_j) & \text{if } j \neq 1 \end{cases}$$

$$= \begin{cases} 0 & \text{if } j = 1 \\ \dfrac{(j-1)(n-i+1)}{n+1}(f_{j-1} - f_j) & \text{if } 2 \leq j \leq i+1 \\ \dfrac{i(n-i+2-k)}{n+1}(f_{j-1} - f_j) & \text{if } j = i+k \ (k = 2, 3, \cdots, n-i) \end{cases}$$

Hence, for any i,

$$\begin{aligned}
\sum_{j=1}^{n} |(\mathbf{A_1}^{-1}\mathbf{F}'')_{i,j}| &\leq \frac{n-i+1}{n+1}(1 + 2 + \cdots + i)|f'| \\
&\quad + \frac{i}{n+1}\left[(n-i) + (n-i-1) + \cdots + 3 + 2\right]|f'|, \\
&\quad \text{where } |f'| = \max|f_{j-1} - f_j| = O(\Delta t \cdot h^2) \\
&\leq \frac{(i+1)i}{2}O(\Delta t \cdot h^2) + \frac{(n-i+2)(n-i-1)}{2}O(\Delta t \cdot h^2) \\
&= O(\Delta t).
\end{aligned}$$

Therefore,

$$\begin{aligned} \|\mathbf{A_1}^{-1}(\mathbf{\Lambda_1 A_1} - \mathbf{A_1 \Lambda_1})\|_\infty &\leq \|(\mathbf{A_1}^{-1}\mathbf{F}')\|_\infty + \|(\mathbf{A_1}^{-1}\mathbf{F}'')\|_\infty \\ &\leq O(\Delta t). \end{aligned}$$

■

Theorem 5.3.3 *Let* $\mathbf{A_1}$ *and* $\mathbf{B_1}$ *be the matrices described in Lemma 5.2.1 and Lemma 5.3.1, respectively. Then*

$$\|\mathbf{C}^{-1}\mathbf{D}\|_\infty \leq 1 + O(\Delta t),$$

where $\mathbf{C} = \mathbf{A_1} - \frac{h}{2}\mathbf{B_1}$ *and* $\mathbf{D} = \mathbf{\Lambda_1 A_1}$.

Proof: Note that

$$\|\mathbf{C}^{-1}\mathbf{D}\|_\infty \leq \bar{\lambda} + \bar{\lambda}\frac{h}{2}\frac{\|\mathbf{A_1}^{-1}\mathbf{B_1}\|_\infty}{1 - \frac{h}{2}\|\mathbf{A_1}^{-1}\mathbf{B_1}\|_\infty} + \frac{\|\mathbf{A_1}^{-1}(\mathbf{\Lambda_1 A_1} - \mathbf{A_1 \Lambda_1})\|_\infty}{1 - \frac{h}{2}\|\mathbf{A_1}^{-1}\mathbf{B_1}\|_\infty}, \qquad (5.10)$$

whenever $\frac{h}{2}\|\mathbf{A_1}^{-1}\mathbf{B_1}\|_\infty < 1$, and where $\bar{\lambda} = \max \lambda_i \leq 1$.

Therefore, from Lemma 5.3.1 and Lemma 5.3.2, we have

$$\begin{aligned} \|\mathbf{C}^{-1}\mathbf{D}\|_\infty &\leq \bar{\lambda} + O(\Delta t) + O(\Delta t) \\ &\leq 1 + O(\Delta t)\ . \end{aligned}$$

■

5.3.2 Max-Norm for Matrix $\mathbf{C}^{-1}\mathbf{D}$ in 2- and 3-Dimensional Spaces

Now we extend the results from the 1-dimensional case to 2- and 3-dimensions. First, we consider the 2-dimensional case. Let $\mathbf{A_2}$ be a tridiagonal block matrix with $(\mathbf{A_1} + 2\mathbf{I})$'s on the diagonal block, and $-\mathbf{I}$'s on the superdiagonal and subdiagonal block. Let $\mathbf{B_2}$ and $\mathbf{\Lambda_2}$ be diagonal block matrices with $\mathbf{B_1}$'s and $\mathbf{\Lambda_1}$'s

on the diagonal block respectively. Then

$$h\mathbf{A_2}^{-1}\mathbf{B_2} =$$
$$h\begin{bmatrix} \mathbf{I}+2\mathbf{A_1}^{-1} & -\mathbf{A_1}^{-1} & & & & \\ -\mathbf{A_1}^{-1} & \mathbf{I}+2\mathbf{A_1}^{-1} & -\mathbf{A_1}^{-1} & & & \\ & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ & & & -\mathbf{A_1}^{-1} & \mathbf{I}+2\mathbf{A_1}^{-1} & -\mathbf{A_1}^{-1} \\ & & & & -\mathbf{A_1}^{-1} & \mathbf{I}+2\mathbf{A_1}^{-1} \end{bmatrix}^{-1}$$
$$\times \begin{bmatrix} \mathbf{A_1}^{-1} & & & \\ & \mathbf{A_1}^{-1} & & \\ & & \ddots & \\ & & & \mathbf{A_1}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{B_1} & & & \\ & \mathbf{B_1} & & \\ & & \ddots & \\ & & & \mathbf{B_1} \end{bmatrix},$$

and we can rewrite it as

$$h\mathbf{A_2}^{-1}\mathbf{B_2} =$$
$$h\mathbf{S_2} \begin{bmatrix} \mathbf{A_1}^{-1} & & & \\ & \mathbf{A_1}^{-1} & & \\ & & \ddots & \\ & & & \mathbf{A_1}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{B_1} & & & \\ & \mathbf{B_1} & & \\ & & \ddots & \\ & & & \mathbf{B_1} \end{bmatrix}, \quad (5.11)$$

where $\mathbf{S_2}$ is the matrix

$$\begin{bmatrix} \mathbf{I}+2\mathbf{A_1}^{-1} & -\mathbf{A_1}^{-1} & & & & \\ -\mathbf{A_1}^{-1} & \mathbf{I}+2\mathbf{A_1}^{-1} & -\mathbf{A_1}^{-1} & & & \\ & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ & & & -\mathbf{A_1}^{-1} & \mathbf{I}+2\mathbf{A_1}^{-1} & -\mathbf{A_1}^{-1} \\ & & & & -\mathbf{A_1}^{-1} & \mathbf{I}+2\mathbf{A_1}^{-1} \end{bmatrix}^{-1}.$$

Thus,

$$h\|\mathbf{A_2}^{-1}\mathbf{B_2}\|_\infty \leq h\|\mathbf{A_1}^{-1}\mathbf{B_1}\|_\infty \|\mathbf{S_2}\|_\infty.$$

By Lemma 5.3.1, we have $\|\mathbf{A_1}^{-1}\mathbf{B_1}\|_\infty = O(\Delta t/h)$. Hence,

$$h\|\mathbf{A_2}^{-1}\mathbf{B_2}\|_\infty \leq O(\Delta t)\|\mathbf{S_2}\|_\infty.$$

Note that $\mathbf{S_2}$ is difficult to analyze using the max-norm. Hence, $\mathbf{S_2}$ is evaluated on a computer to get a numerical value for $\|\mathbf{S_2}\|_\infty$. For example, if $\mathbf{A_1}$ is 10×10 and $m = 10$, so that $\mathbf{S_2}$ is 100×100, then $\|\mathbf{S_2}\|_\infty \approx 1.6687$. If $\mathbf{A_1}$ is 20×20 and $m = 20$, so that $\mathbf{S_2}$ is 400×400, then $\|\mathbf{S_2}\|_\infty \approx 2.0794$. If $\mathbf{A_1}$ is 40×40 and $m = 40$, so that $\mathbf{S_2}$ is 1600×1600, then $\|\mathbf{S_2}\|_\infty \approx 2.5041$. Therefore, we have

$$h\|\mathbf{A_2}^{-1}\mathbf{B_2}\|_\infty = O(\Delta t). \tag{5.12}$$

Also, we have

$$
\begin{aligned}
&\mathbf{A_2}^{-1}(\mathbf{\Lambda_2 A_2 - A_2 \Lambda_2}) \\
&= \begin{bmatrix} \mathbf{I}+2\mathbf{A_1}^{-1} & -\mathbf{A_1}^{-1} & & & \\ -\mathbf{A_1}^{-1} & \mathbf{I}+2\mathbf{A_1}^{-1} & -\mathbf{A_1}^{-1} & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \ddots \\ & & & -\mathbf{A_1}^{-1} & \mathbf{I}+2\mathbf{A_1}^{-1} \end{bmatrix}^{-1} \\
&\quad \times \begin{bmatrix} \mathbf{A_1}^{-1} & & \\ & \ddots & \\ & & \mathbf{A_1}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda_1 A_1 - A_1 \Lambda_1} & & \\ & \ddots & \\ & & \mathbf{\Lambda_1 A_1 - A_1 \Lambda_1} \end{bmatrix} \\
&= \mathbf{S_2} \begin{bmatrix} \mathbf{A_1}^{-1} & & \\ & \ddots & \\ & & \mathbf{A_1}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda_1 A_1 - A_1 \Lambda_1} & & \\ & \ddots & \\ & & \mathbf{\Lambda_1 A_1 - A_1 \Lambda_1} \end{bmatrix},
\end{aligned}
$$

and then

$$
\begin{aligned}
\|\mathbf{A_2}^{-1}(\mathbf{\Lambda_2 A_2 - A_2 \Lambda_2})\|_\infty &\leq \|\mathbf{A_1}^{-1}(\mathbf{\Lambda_1 A_1 - A_1 \Lambda_1})\|_\infty \|\mathbf{S_2}\|_\infty \\
&= O(\Delta t), \qquad (5.13)
\end{aligned}
$$

since $\|\mathbf{A_1}^{-1}(\mathbf{\Lambda_1 A_1 - A_1 \Lambda_1})\|_\infty = O(\Delta t)$ by Lemma 5.3.2. From Equation (5.10), we have the estimate $\|\mathbf{C}^{-1}\mathbf{D}\|_\infty = 1 + O(\Delta t)$.

Three dimensions can be analyzed in the same manner. We write $\mathbf{A}$ as a tridiagonal block matrix with $(\mathbf{A_2} + 2\mathbf{I})$'s on the diagonal block and with $-\mathbf{I}$'s on the superdiagonal and subdiagonal block. Write $\mathbf{B}$ and $\mathbf{\Lambda}$ as diagonal matrices with

$\mathbf{B_2}$'s and $\mathbf{\Lambda_2}$'s on the diagonal block respectively. Write

$$h\mathbf{A}^{-1}\mathbf{B} = h\mathbf{S_3} \begin{bmatrix} \mathbf{A_2}^{-1} & & & \\ & \mathbf{A_2}^{-1} & & \\ & & \ddots & \\ & & & \mathbf{A_2}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{B_2} & & & \\ & \mathbf{B_2} & & \\ & & \ddots & \\ & & & \mathbf{B_2} \end{bmatrix},$$

and

$$\mathbf{A}^{-1}(\mathbf{\Lambda A} - \mathbf{A\Lambda}) = \mathbf{S_3} \begin{bmatrix} \mathbf{A_2}^{-1} & & \\ & \ddots & \\ & & \mathbf{A_2}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda_2 A_2} - \mathbf{A_2 \Lambda_2} & & \\ & \ddots & \\ & & \mathbf{\Lambda_2 A_2} - \mathbf{A_2 \Lambda_2} \end{bmatrix},$$

where the matrix $\mathbf{S_3}$ is

$$\begin{bmatrix} \mathbf{I} + 2\mathbf{A_2^{-1}} & -\mathbf{A_2^{-1}} & & & & \\ -\mathbf{A_2^{-1}} & \mathbf{I} + 2\mathbf{A_2^{-1}} & -\mathbf{A_2^{-1}} & & & \\ & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ & & & -\mathbf{A_2^{-1}} & \mathbf{I} + 2\mathbf{A_2^{-1}} & -\mathbf{A_2^{-1}} \\ & & & & -\mathbf{A_2^{-1}} & \mathbf{I} + 2\mathbf{A_2^{-1}} \end{bmatrix}^{-1}.$$

To discuss the max-norm, we still need to evaluate the factor $\mathbf{S_3}$. Using a computer to evaluate the max-norm of $\mathbf{S_3}$, we have: if $\mathbf{A_2}$ is 5×5, so that $\mathbf{S_3}$ is 125×125, then $\|\mathbf{S_3}\|_\infty \approx 1.9841$. If $\mathbf{A_2}$ is 10×10, so that $\mathbf{S_3}$ is 1000×1000, then $\|\mathbf{S_3}\|_\infty \approx$

2.5126. Similarly, we have

$$\begin{aligned} h\|\mathbf{A}^{-1}\mathbf{B}\|_\infty &\leq h\|\mathbf{A_2}^{-1}\mathbf{B_2}\|_\infty\|\mathbf{S_3}\|_\infty \\ &= O(\Delta t) \end{aligned}$$

and

$$\begin{aligned} \|\mathbf{A}^{-1}(\mathbf{\Lambda A} - \mathbf{A\Lambda})\|_\infty &\leq \|\mathbf{A_2}^{-1}(\mathbf{\Lambda_2 A_2} - \mathbf{A_2\Lambda_2})\|_\infty\|\mathbf{S_3}\|_\infty \\ &= O(\Delta t) \end{aligned}$$

by Equation (5.12) and Equation (5.13) from the 2-dimensional space. Finally, we have $\|\mathbf{C}^{-1}\mathbf{D}\|_\infty = 1 + O(\Delta t)$ from Equation (5.10).

5.3.3 Numerical Results for Spectral Radius of $\mathbf{C}^{-1}\mathbf{D}$

Next we use the approximate value of γ (Section 3.3) to calculate some numerical results for spectral radius of $\mathbf{C}^{-1}\mathbf{D}$, (Table 5-1, Table 5-2 and Table 5-3.), in the domain Ω_1 when $\Delta t = 0.1$ s. Numerically, we observe that the spectral radius of $\mathbf{C}^{-1}\mathbf{D}$ is strictly less than 1.

Table 5–1: Spectral radius for $\mathbf{C}^{-1}\mathbf{D}$ in one dimensional space

Matrix dimension ($n \times n$)	Spectral radius for $\mathbf{C}^{-1}\mathbf{D}$
$n = 50$	0.9990
$n = 100$	0.9991
$n = 200$	0.9992
$n = 400$	0.9992
$n = 800$	0.9992
$n = 1600$	0.9992

Table 5–2: Spectral radius for $\mathbf{C}^{-1}\mathbf{D}$ in two dimensional space

Matrix dimension $(n \times n)$	Spectral radius for $\mathbf{C}^{-1}\mathbf{D}$
$n = 100$ $(m = 10)$	0.9960
$n = 400$ $(m = 20)$	0.9984
$n = 1600$ $(m = 40)$	0.9989
$n = 6400$ $(m = 80)$	0.9991

Table 5–3: Spectral radius for $\mathbf{C}^{-1}\mathbf{D}$ in three dimensional space

Matrix dimension $(n \times n)$	Spectral radius for $\mathbf{C}^{-1}\mathbf{D}$
$n = 125$ $(m = 5)$	0.9715
$n = 1000$ $(m = 10)$	0.9960
$n = 8000$ $(m = 20)$	0.9984
$n = 27000$ $(m = 30)$	0.9988

CHAPTER 6
NUMERICAL EXPERIMENTS

This chapter briefly reports some numerical experiments by using the finite difference scheme described in Chapter 3. In these applications, the FORTRAN package FISHPAK (see Appendix B) is utilized to deal with large sparse matrices in our model and to solve linear separable elliptic equations in three space dimensions.

The conductivity function is computed as

$$\sigma = 10^{p(z)}, \tag{6.1}$$

where p(z) is the quartic polynomial (see Equation (3.21) and Table 3-1) and the altitude z is measured in kilometers. Let us assume as in Hager et al. [7] that the current per unit volume at position $\mathbf{r}$ is

$$\nabla \cdot \mathbf{J} = A_P e^{-\omega|\mathbf{r}-\mathbf{s_P}|^2} - A_N e^{-\omega|\mathbf{r}-\mathbf{s_N}|^2},$$

where $\mathbf{s_P}$ and $\mathbf{s_N}$ are the centers of the positively charged and negatively charged current generators, respectively. The parameters A_P and A_N are chosen so that the total current generated by either the positive or the negative charge centers has magnitude 1 A. Hence, we have

$$A_P \int_{z \geq -h_1} e^{-\omega R^2} dV = 1,$$

where h_1 is the altitude for the positive charge center, $R^2 = |\mathbf{r} - \mathbf{s_P}|^2$, and z is the altitude measured with respect to the positive charge center. The cylindrical shell method is applied to the above integral, as shown in Figure 6–1, to obtain

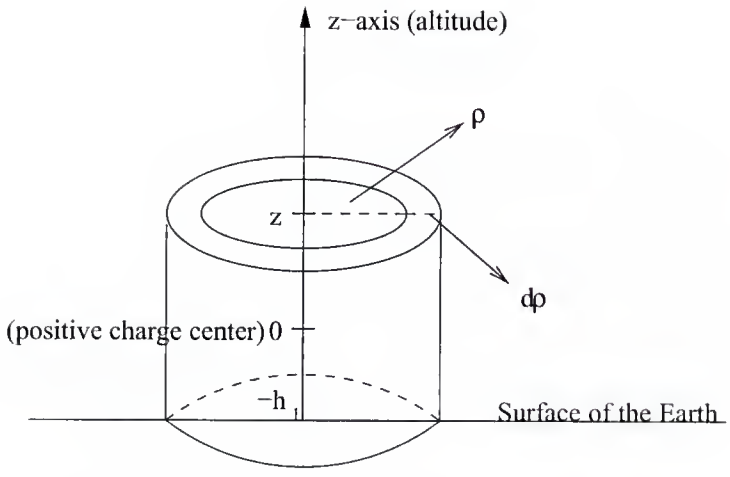

Figure 6–1: Cylindrical shell of the volume integral $\int dV$

$$\begin{aligned}
\int_{z\geq -h_1} e^{-\omega R^2} dV &= \int_{z\geq -h_1\ ,\ \rho\geq 0} e^{-\omega(\rho^2+z^2)} 2\pi\rho dz d\rho \\
&= \left(\int_{\rho\geq 0} 2\pi\rho e^{-\omega\rho^2} d\rho\right)\left(\int_{z\geq -h_1} e^{-\omega z^2} dz\right) \\
&= \left(\frac{\pi}{\omega}\right)\left[\frac{1}{2}\left(\frac{\pi}{\omega}\right)^{1/2} + \frac{1}{2}\left(\frac{\pi}{\omega}\right)^{1/2} erf(\sqrt{\omega}h_1)\right] \\
&= \left(\frac{\pi}{\omega}\right)^{3/2} \frac{[1+erf(\sqrt{\omega}h_1)]}{2}
\end{aligned}$$

where erf is the error function defined as

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt.$$

As a result, A_P is calculated as

$$A_P = \left(\frac{\omega}{\pi}\right)^{3/2} \frac{2}{[1+erf(\sqrt{\omega}h_1)]},$$

and, similarly, A_N can be obtained

$$A_N = \left(\frac{\omega}{\pi}\right)^{3/2} \frac{2}{[1+erf(\sqrt{\omega}h_2)]},$$

where h_2 is the altitude for the negative charge center. Therefore,

$$\nabla \cdot \mathbf{J} = \left(\frac{\omega}{\pi}\right)^{3/2} \left(\frac{2}{[1 + erf(\sqrt{\omega} h_1)]} e^{-\omega |\mathbf{r} - \mathbf{s_P}|^2} - \frac{2}{[1 + erf(\sqrt{\omega} h_2)]} e^{-\omega |\mathbf{r} - \mathbf{s_N}|^2} \right)$$

is used in our model. In each experiment, the initial potential of the entire domain is assumed to be zero. The centers of the current generators are positioned on the $z-$axis, at altitudes 5 km (-) and 10 km (+).

Four experiments are conducted in this chapter. Section 6.1 uses different values of ω to find the breakdown time with the ambient conductivity and the reduced conductivity. Section 6.2 explores a finer mesh for our model with an acceptable accuracy for the potential ϕ. Section 6.3 and Section 6.4 check the temporal error and the spatial error numerically respectively.

6.1 Experiment 1

This experiment considers the domain Ω_1, previously introduced:

$$\Omega_1 = \{(x, y, z) | 0 \leq z \leq 100 \text{ km}, \ -50 \text{ km} \leq x, y \leq 50 \text{ km}\}$$

For conductivity, we use the ambient conductivity given in Equation (6.1). For different values of ω, we integrate the Equation (3.7) solving for potential ϕ. The integration is stopped when the electric field reaches the breakdown threshold, which is taken to be 250,000 V m^{-1} [7]. We use the uniform mesh in all $x-$, $y-$, and $z-$directions with mesh size 1000 m, and the time step is $\Delta t = 1/16$ s. As shown in Table 6-1, the breakdown time is calculated based on different values of ω.

Table 6-1 demonstrates that the breakdown time shortens rapidly as ω increases. For example, when $\omega = 4$, the breakdown occurs within 15.5625 seconds. With $\omega = 16$, the breakdown time drops to 1.1850 seconds. This result is supported by [7]: as ω increases, the current becomes more concentrated at $\mathbf{s_P}$ and $\mathbf{s_N}$.

Observe that when $\omega = 1$, there is no breakdown within one hour. According to our numerical results, the maximum electric field approaches the limit 232429.88

Table 6–1: Breakdown time comparison for different values of ω

ω value	breakdown time with the ambient conductivity (without cloud)	breakdown time with the reduced conductivity (with cloud)
$\omega = 1$	∞	64.7500 s
$\omega = 4$	15.5625 s	13.6875 s
$\omega = 16$	1.1850 s	1.75000 s

V/m , which is less than the breakdown threshold 250,000 V/m, as t tends to infinity. Therefore, the breakdown never happens when $\omega = 1$.

In a thundercloud, the conductivity is less than the ambient conductivity that appears in Equation (6.1). Next, we consider the effect of reducing conductivity in order to simulate a cloud. More precisely, the ambient conductivity is multiplied by the factor 0.05 as described in [7]. The results are shown in the third column of Table 6-1. Again, the breakdown time decreases dramatically as ω increases. This pattern is consistent with the previous computed breakdown times for the ambient conductivity.

As can be seen in Table 6-1, reducing in conductivity to simulate a cloud leads to the occurrence of breakdown when $\omega = 1$, while with the ambient conductivity, charge is conducted into the surrounding atmosphere faster enough to prevent the electric field from reaching the breakdown threshold. In the remaining experiments, the conductivity is given by Equation (6.1) times 0.05 (to simulate a cloud).

6.2 Experiment 2

A finer mesh in our model is expected to obtain more accurate results. However, such refinement is not only time consuming, but also increasing memory usage. In order to maintain the accuracy of our results efficiently, a graded mesh is considered. We place a large percentage of grid points near the charge centers along the $z-$ direction where the potential changes rapidly. The grid along the

$z-$axis is chosen in the following manner [7]:

$$z_k = k\tau \qquad \text{for } 1 \leq k \leq \frac{n}{2} \tag{6.2}$$

and

$$z_k = 2H + \tau s^{\left(k-\frac{n}{2}\right)} \qquad \text{for } \frac{n}{2} < k \leq n, \tag{6.3}$$

where H is the altitude of the positive charge center (10 km), $\tau = 4H/n$, and s is chosen so that $2H + \tau s^{(n/2)} = 100$ km is the height of the domain. According to Hager et al. [7], the vicinity of the charge centers is critical to efficiency. Therefore, the distribution of the grid points is uniform below 2H and geometric from 2H to 100 km. We still use the uniform mesh for the $x-$ and $y-$directions as in Experiment 1.

Let us consider two domains: Ω_1 as described in Section 5.1 and Ω_2 as follows,

$$\Omega_2 = \{(x, y, z) | 0 \leq z \leq 100 \text{ km}, \ -25 \text{ km} \leq x, y \leq 25 \text{ km}\}.$$

In this experiment, the potentials are found at the time 65 seconds with the time step $\Delta t = 1/2$ s and the ambient conductivity reduced by the factor 0.05 when $\omega = 1$. The graded mesh is applied with 200 grid points along $z-$axis (mesh size = 200 m for $0 \leq z \leq 20$ km). The mesh size is 1000 m for both in $x-$ and $y-$directions. Figure 6–2 shows the potentials in two domains, Ω_1 and Ω_2, for $x = 0$, $y = 0$, and $0 \leq z \leq 100$ km. The potential from Ω_1 is shown in a solid line and that from Ω_2 is shown in a dashed line. The absolute difference in potential is about 10^5 and the relative difference is about 10^{-4} between these two domains. Therefore, we cannot see any difference for the potentials in Figure 6–2. Figure 6–3 and Figure 6–4 are the potentials near the negative and positive charge centers respectively, along the $z-$axis when $x = 0$ and $y = 0$.

Since the relative difference for potentials is quite small between these two domains, we only focus on the smaller domain Ω_2 for the remaining experiments.

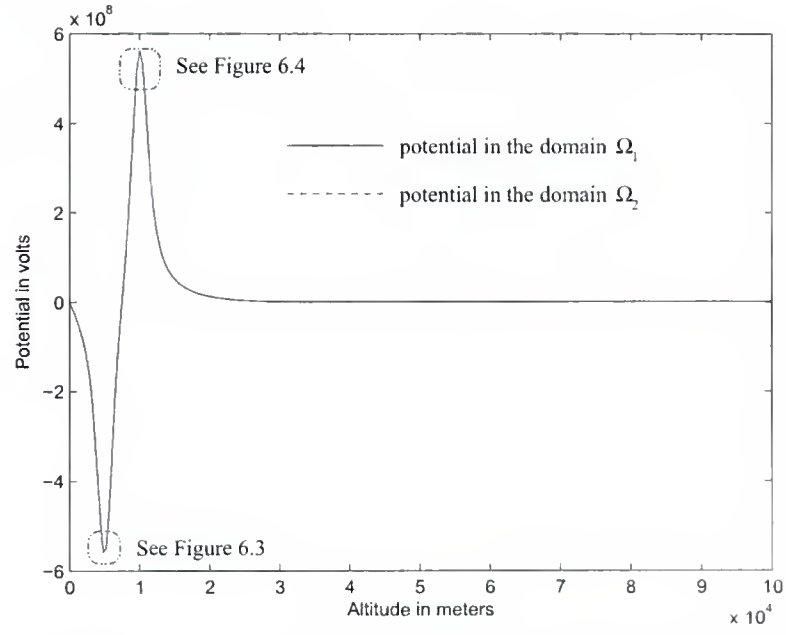

Figure 6–2: Potentials ($\phi(0,0,z), 0 \leq z \leq 100$ km) in two domains, Ω_1 and Ω_2

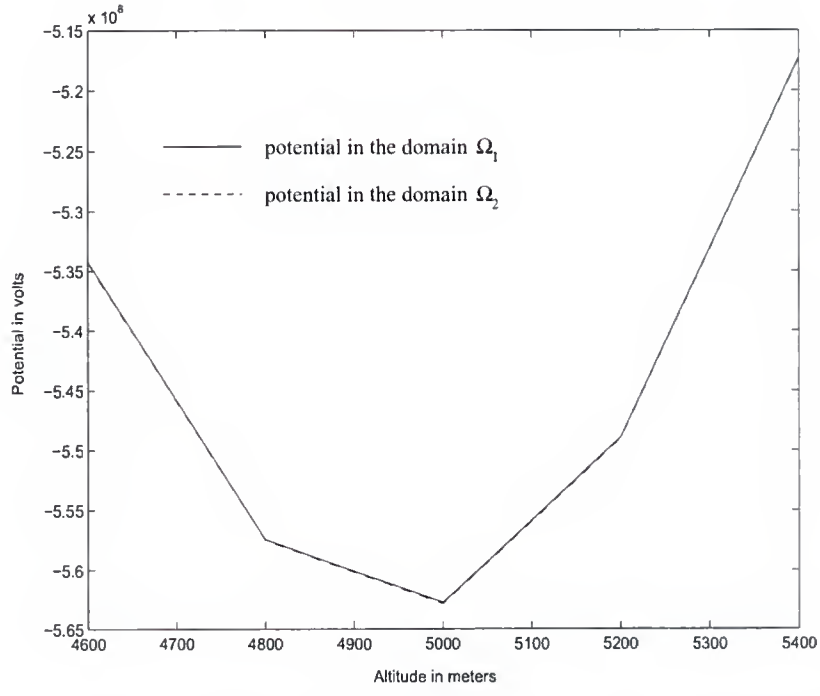

Figure 6–3: Potentials ($\phi(0,0,z)$, 4.6 km $\leq z \leq$ 5.4 km) near the negative charge (5 km) in two domains, Ω_1 and Ω_2

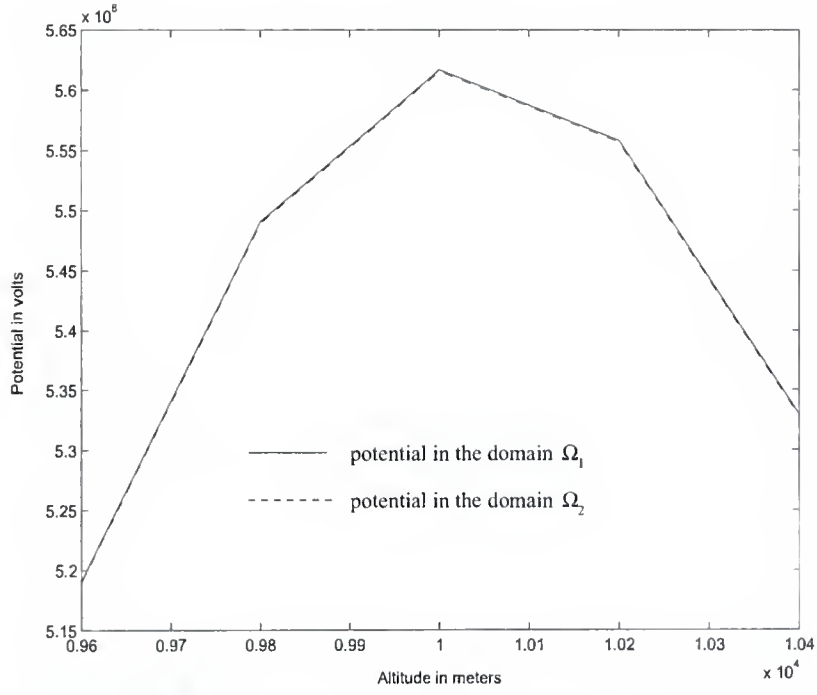

Figure 6–4: Potentials ($\phi(0, 0, z)$, 9.6 km $\leq \ z \ \leq$ 10.4 km) near the positive charge (10 km) in two domains, Ω_1 and Ω_2

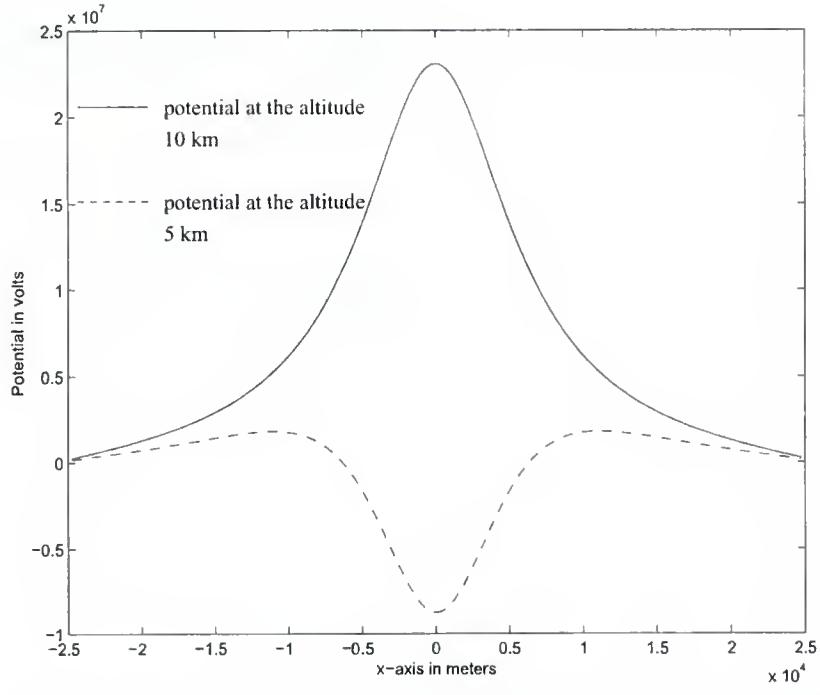

Figure 6–5: Potentials ($\phi(x, 0, 5)$ and $\phi(x, 0, 10)$, -25 km $\leq \ x \ \leq$ 25 km) along the $x-$axis at the altitude 5 km and 10 km

Further, we find the potentials (see Figure 6–5) along the $x-$axis at the altitude 5 km (the negative charge center) and 10 km (the positive charge center) when $y = 0$. In Figure 6–5, the solid line represents the potential at the altitude 10 km and the dashed line represents the potential at the altitude 5 km.

6.3 Experiment 3

According to the error estimate described in Theorem 4.2.4, when $\mu = 1/2$ (Crank-Nicholson scheme), the temporal error term should be smaller than for other values of μ. To verify the temporal error analysis numerically, the potential is estimated at time t = 65 s, which is close to the breakdown time 64.75 seconds when $\omega = 1$. The graded mesh (see Equation (6.2) and Equation(6.3)) is applied with 200 grid points along z-axis (mesh size = 200 m for $0 \leq z \leq 20$ km). The mesh size is 1000 m for both in x- and y- axes. We compare the errors in different time steps (1/2, 1/4, $\cdots$, 1/64) for $\mu = 1/2$ and $\mu = 1$. To analyze the error numerically, we use the maximum norm:

$$\|\mathbf{\Phi}_{\Delta t}\|_\infty = \max_{i,j,k} |(\phi_{\Delta t})_{i,j,k}|,$$

where $(\phi_{\Delta t})_{i,j,k}$ is the computed potential using the time step Δt at each mesh point in the domain at t = 65 s, and $\mathbf{\Phi}_{\Delta t}$ is the potential vector with elements $(\phi_{\Delta t})_{i,j,k}$. First, we consider the potential corresponding to the time step $\Delta t = 1/128$ s as the "exact" potential and subtract it from the potentials corresponding to the compared time steps to obtain error estimates at all mesh points. The maximum norm error at different time step is shown in the third column of Table 6-2 and Table 6-3. Second, the relative errors are estimated by the ratio

$$\frac{\|\mathbf{\Phi}_{\Delta t} - \mathbf{\Phi}_{1/128}\|_\infty}{\|\mathbf{\Phi}_{1/128}\|_\infty},$$

as shown in the fourth column of Table 6-2 and Table 6-3.

Table 6–2: Relative error in the potential at $t = 65$ seconds using the maximum norm for Crank-Nicholson scheme ($\mu = 1/2$).

Time step (s)	maximum norm potential	maximum norm error with respect to time step 1/128 s	Crank-Nicholson relative error ($\mu = 1/2$)
1/2	564431185.91904	80.21113	1.42110×10^{-7}
1/4	564431163.61478	20.30029	3.59659×10^{-8}
1/8	564431157.94013	5.11413	9.06068×10^{-9}
1/16	564431156.50102	1.27381	2.25681×10^{-9}
1/32	564431156.13736	0.30471	5.39852×10^{-10}
1/64	564431156.04677	0.06052	1.07192×10^{-10}
1/128	564431156.04677	0.00000	0.00000

Table 6–3: Relative error in the potential at $t = 65$ seconds using the maximum norm for Backward Euler scheme ($\mu = 1$).

Time step (s)	maximum norm potential	maximum norm error with respect to time step 1/128 s	Backward Euler relative error ($\mu = 1$)
1/2	564422403.52880	35278.79275	6.25033×10^{-5}
1/4	564426772.69223	17375.01373	3.07832×10^{-5}
1/8	564428962.54654	8411.09363	1.49019×10^{-5}
1/16	564430058.82113	3926.08120	6.95582×10^{-6}
1/32	564430607.30153	1682.80248	2.98141×10^{-6}
1/64	564430881.63001	560.96757	9.93864×10^{-7}
1/128	564430881.63001	0.00000	0.00000

In Table 6-2 which corresponds to the Crank-Nicholson scheme, if the time step Δt is multiplied by $1/2$, the relative error is multiplied by $1/4$. Hence, the error for Crank-Nicholson scheme ($\mu = 1/2$) behaves like $O(\Delta t)^2$. In Table 6-3 which corresponds to the backward Euler difference scheme, if the time step Δt is multiplied by $1/2$, the relative error is also multiplied by $1/2$. Therefore, the error for backward Euler difference scheme ($\mu = 1$) behaves like $O(\Delta t)$. This suggests that the Crank-Nicholson scheme is far superior to the backward Euler difference scheme.

Note that the forward difference scheme ($\mu = 0$) does not work, since it is not stable as we described in Section 4.1. Experimentally, taking $\mu = 0$, we find that the max-norm of the potential grows exponentially in t. The plot of $\log_{10}(\|\Phi_{\Delta t}\|_\infty)$ versus t appears in Figure 6–6, with time step $\Delta t = 1/8$ s. Observe that the maximum potential at $t = 65$ seconds is 1.04737×10^{89}.

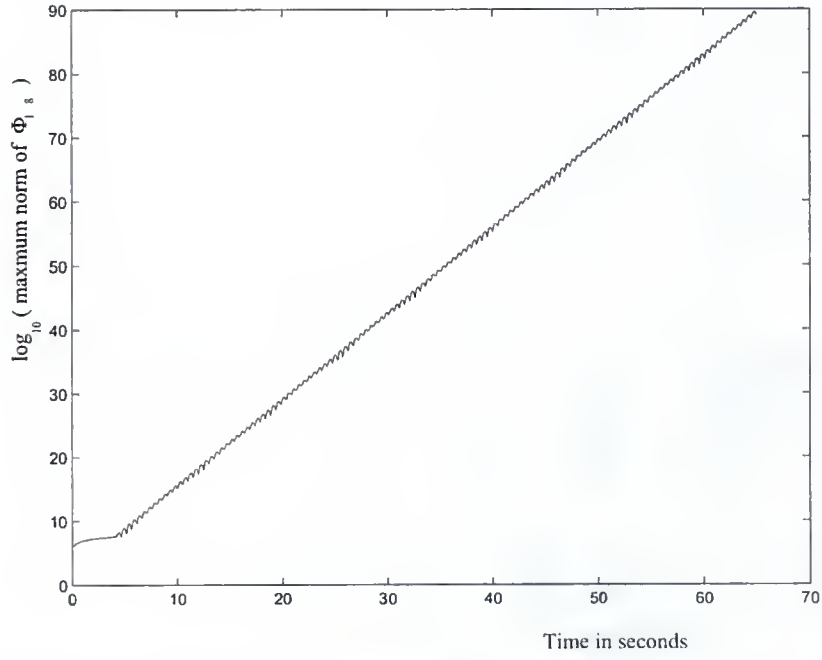

Figure 6–6: Maximum norm of potential versus time, where $\Delta t = 1/8$ s and $\mu = 0$

6.4 Experiment 4

The last study is to check the spatial error term. We take the time step $\Delta t = 1/2$ s and use Crank-Nicholson scheme $\mu = 1/2$. Four different graded meshes are used in the domain Ω_2, where

$$\Omega_2 = \{(x, y, z) | 0 \leq z \leq 100 \text{ km}, \ -25 \text{ km} \leq x, y \leq 25 \text{ km}\},$$

and the grid points are chosen according to Equation (6.2) and Equation (6.3) along the $z-$axis:

- Case 1: 30 uniform grid points in the $x-$ and $y-$directions, and 100 grid points in the $z-$ direction (mesh size = 400 m for $0 \leq z \leq 20$ km, and exponentially increasing mesh for $20 \leq z \leq 100$ km, from $\Delta z = 400$ m near $z = 20$ km to $\Delta z = 8043.61$ m near $z = 100$ km).
- Case 2: 60 uniform grid points in the $x-$ and $y-$directions, and 200 grid points in the $z-$ direction (mesh size = 200 m for $0 \leq z \leq 20$ km, and exponentially increasing mesh for $20 \leq z \leq 100$ km, from $\Delta z = 200$ m near $z = 20$ km to $\Delta z = 4652.41$ m near $z = 100$ km).
- Case 3: 120 uniform grid points in the $x-$ and $y-$directions, and 400 grid points in the $z-$ direction (mesh size = 100 m for $0 \leq z \leq 20$ km, and exponentially increasing mesh for $20 \leq z \leq 100$ km, from $\Delta z = 100$ m near $z = 20$ km to $\Delta z = 2629.65$ m near $z = 100$ km).
- Case 4: 240 uniform grid points in the $x-$ and $y-$directions, and 800 grid points in the $z-$ direction (mesh size = 50 m for $0 \leq z \leq 20$ km, and exponentially increasing mesh for $20 \leq z \leq 100$ km, from $\Delta z = 50$ m near $z = 20$ km to $\Delta z = 1462.03$ m near $z = 100$ km).

Let ϕ^n, ϕ^{2n}, ϕ^{4n}, and ϕ^{8n} denote the discrete potential evaluated along z-axis at $z = 10$ km, where n, $2n$, $4n$, and $8n$ are the numbers of mesh intervals along $z-$axis in case 1, 2, 3, and 4 respectively. We focus on $z = 10$ km since the

the largest potential is achieved here. Suppose that ϕ^{∞} is the "exact" potential at $z = 10$ km, and the sequence ϕ^{n}, ϕ^{2n}, ϕ^{4n}, $\cdots$ is converging to the limit ϕ^{∞}. Extrapolation techniques (see [3], Section1.8) are used to estimate various constants in the expression of the error estimate.

First, based on Theorem 4.2.4, the spatial error is $O(h^2)$. Hence,

$$\phi^{n} = \phi^{\infty} + E_n, \tag{6.4}$$

where E_n is the error term with the property $|E_n| \leq Ch^2$, and C is a constant, and $h = 1/n$. In extrapolation techniques, we treat the inequality $|E_n| \leq Ch^2$ as an equality. Two elements in the sequence, say ϕ^{n} and ϕ^{2n}, are used in order to solve the constant C. Plugging ϕ^{n} and ϕ^{2n} into the error relation Equation (6.4), we have

$$\phi^{n} - \phi^{\infty} = E_n, \tag{6.5}$$

$$\phi^{2n} - \phi^{\infty} = E_{2n}. \tag{6.6}$$

Subtract Equation (6.6) from Equation (6.5) to obtain

$$\begin{aligned} \phi^{n} - \phi^{2n} &= E_n - E_{2n} \\ &= Ch^2 - C\left(\frac{h}{2}\right)^2 \\ &= \frac{3}{4}Ch^2. \end{aligned}$$

Therefore,

$$C = \frac{\phi^{n} - \phi^{2n}}{\frac{3}{4}h^2},$$

where $h = 1/n$.

Any two potential vectors ϕ^i and ϕ^{2i} can be employed in the same way as above to find the constant C:

$$C = \frac{\phi^i - \phi^{2i}}{\frac{3}{4i^2}}. \tag{6.7}$$

Table 6–4: Estimates for the constant C generated by Equation (6.7)

n	$\phi^n - \phi^{2n}$	C
100		
	175594927.29868	2.34×10^8
200		
	25121434.49608	1.34×10^8
400		
	5574244.17420	1.19×10^8
800		

In Table 6-4, only four different meshes are considered. If $n < 100$, the potential vector is not accurate enough, and if $n > 800$, there is not enough memory to do computations on our computer network. The maximum norms of the difference of two successive potential vectors are shown in the second column of Table 6-4. Observe that the maximum norm of the difference are getting smaller. We can predict that it approaches 0 when n tends to infinity if more data (potentials for larger n) provided. This shows that the sequence ϕ^n is converging. The estimates for the constant C are shown in the third column of Table 6-4. The relative difference for the last two estimates of C is 0.13.

Next, let us assume that the spatial error is $O(h^p)$, that is,

$$\phi^n = \phi^\infty + E_n,$$

where $|E_n| \leq Ch^p$, C and p are constants. Using the similar analysis as above, we will compute the constants p and C. For three elements ϕ^n, ϕ^{2n}, and ϕ^{4n}, we have

$$\phi^n - \phi^{2n} = Ch^p\left(1 - \frac{1}{2^p}\right), \tag{6.8}$$

$$\begin{aligned}\phi^{2n} - \phi^{4n} &= Ch^p\left(\frac{1}{2^p} - \frac{1}{4^p}\right) \\ &= \frac{Ch^p}{2^p}\left(1 - \frac{1}{2^p}\right).\end{aligned} \tag{6.9}$$

Equation (6.8) divides Equation(6.9) to get

$$2^p = \frac{\phi^n - \phi^{2n}}{\phi^{2n} - \phi^{4n}}.$$

Therefore,

$$p = \log_2\left(\frac{\phi^n - \phi^{2n}}{\phi^{2n} - \phi^{4n}}\right),$$

and

$$C = \frac{\phi^n - \phi^{2n}}{h^p\left(1 - \dfrac{\phi^{2n} - \phi^{4n}}{\phi^n - \phi^{2n}}\right)}$$

can be found, where $h = 1/n$.

Given any three elements ϕ^i, ϕ^j, and ϕ^k with the property $j^2 = ik$, the formula for constants p and C are:

$$p = \log_2\left(\frac{\phi^i - \phi^j}{\phi^j - \phi^k}\right), \quad C = \frac{i^p(\phi^i - \phi^j)}{1 - \dfrac{\phi^j - \phi^k}{\phi^i - \phi^j}}. \tag{6.10}$$

The computed constants p and C are shown in the third and fourth columns of Table 6-5 respectively. The best estimate for the constant C is 1.45×10^8 from Table 6-5, while the best estimate is 1.19×10^8 from Table 6-4. The relative difference between them is 0.22. From the third column of Table 6-5, we can see that the constant p is getting close to the theoretical value 2.

Table 6–5: Estimates for constants p and C generated by Equation (6.10)

n	$\phi^n - \phi^{2n}$	p	C
100	175594927.29868		
200		2.81	2.04×10^8
400	25121434.49608	2.17	1.45×10^8
800	5574244.17420		

CHAPTER 7
CONCLUSION

In this dissertation, Maxwell's equations are used to establish an electric field model between the surface of the Earth and the ionosphere:

$$\begin{aligned}
\text{PDE:} \quad & \frac{\partial}{\partial t}\triangle\phi + \frac{\sigma}{\varepsilon}\triangle\phi + \left(\frac{\sigma_z}{\varepsilon}\frac{\partial\phi}{\partial z} + \frac{\nabla\cdot\mathbf{J}}{\varepsilon}\right) = 0 \qquad (7.1)\\
\text{BC:} \quad & \phi = 0, \text{ if } z = 0\\
& \phi = 250,000, \text{ if } z = D_z\\
& \phi = \int_0^z \frac{c}{\sigma(z)}dz, \text{ if } x = \pm D_x \text{ or } y = \pm D_y\\
\text{Domain:} \quad & \Omega = \{(x,y,z) : |x| \le D_x, |y| \le D_y, 0 \le z \le D_z\}
\end{aligned}$$

The conductivity function σ varies by many orders of magnitude between the Earth and the ionosphere, which causes the partial differential equation (7.1) to be stiff. We develop stable finite difference approximations to our model that take into account the σ variation. Error estimates are obtained by using the max-norm. Four numerical experiments are conducted to verify our theoretical analysis as described in Chapter 4 and Chapter 5.

APPENDIX A
APPROXIMATION FOR CONDUCTIVITY FUNCTION

MATLAB programs are used to obtain the approximation for the conductivity function. The detailed approximation for the conductivity function from *United Sates Air Force Cambridge Research Laboratories* [27] is given in Appendix A.1. The approximation from *Lightning: physics and effects* [21] is discussed in Appendix A.2.

A.1 Approximation from Figure 9.6 [27]

Four points (see Table A-1) on the curve B from Figure 9.6 [27] are chosen to find an approximation for conductivity. The conductivity graph in Figure 9.6 [27] contains the data up to the altitude 27.828 km (about 100 ft), while our model domain is 100 km. All four points are chosen to have a cubic fit for the altitude $0 \leq z \leq 27.828$ km, and the approximation can be obtained by MATLAB as follows:

$$\sigma(z) = 10^{-13.40+0.1600z-0.005175z^2+0.00009077z^3}.$$

For altitude $28.828 \leq z \leq 100$ km, the last three points (P_2, P_3, and P_4) are picked to have a linear fit, and the approximation is

$$\sigma(z) = 10^{-12.88+0.06718z}.$$

Figure A–1 shows the graph of the approximation for the conductivity function.

A.2 Approximation from Figure 3–3

Thirteen points (see Table A-2) in Figure 3–3 are picked to approximate the conductivity function. The conductivity is chosen as the average of shaded region in Figure 3–3 for the altitude above 30 km. The advantage for using this figure

Table A–1: Four points chosen from Figure 9.6 [27]

Point	Height (km)	$\log_{10}$ (conductivity) (1/ohm-m)
P_1	2.7432	-13
P_2	12.954	-12
P_3	20.7264	-11.5
P_4	27.82824	-11

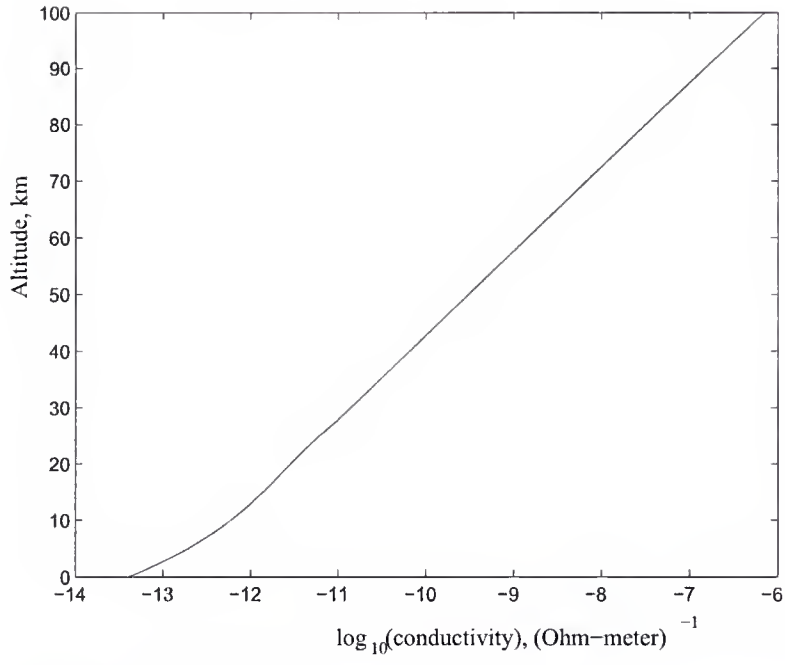

Figure A–1: Electric conductivity as a function of altitude approximated from curve B in Figure 9.6 [27]

to estimate the conductivity is that Figure 3–3 demonstrates a more completed conductivity data than Figure 9.6 in [27].

Table A–2: Thirteen points chosen from Figure 3–3

Point	Height (km)	$\log_{10}$ (conductivity) (S/m)
P_1	4.5	-13
P_2	12.5	-12
P_3	26	-11
P_4	40	-10
P_5	48	-9
P_6	52.5	-8
P_7	60	-7
P_8	70	-6
P_9	78.5	-5
P_{10}	85	-4
P_{11}	89	-3
P_{12}	95	-2
P_{13}	100	-1

A quartic polynomial is found to approximate those points in Table 2. The conductivity function is estimated by

$$\sigma = 10^{p(z)},$$

where p(z) is the quartic polynomial

$$\begin{aligned} p(z) &= -13.16 + 0.06351z + 0.0008682z^2 - 0.000008674z^3 \\ &\quad +0.00000005781z^4, \end{aligned}$$

where the altitude z is measured in kilometers, and its graph is shown in Figure A–2.

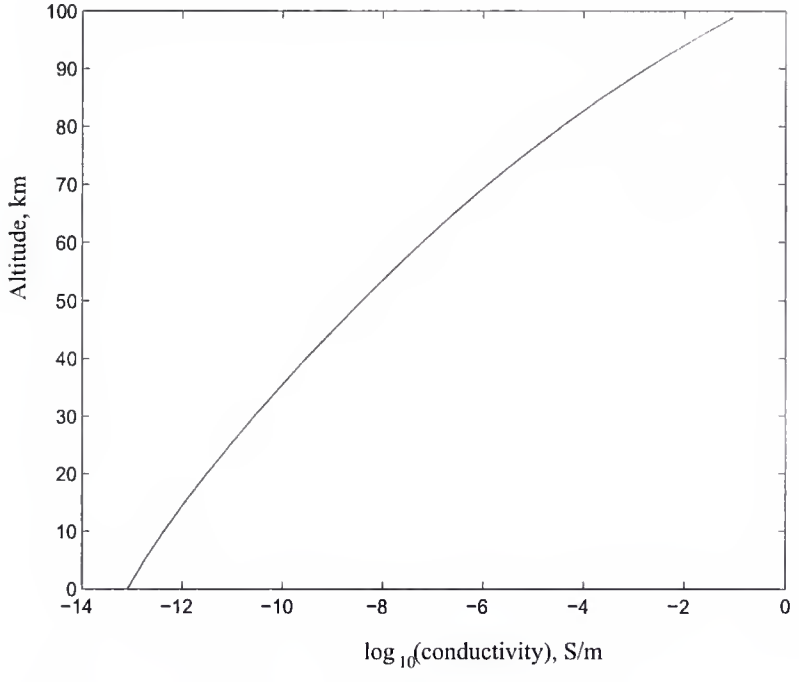

Figure A–2: Electric conductivity as a function of altitude approximated from Figure 3–3

APPENDIX B
FISHPAK

The FORTRAN package FISHPAK developed by Swarztrauber and Sweet is used for our numerical experiments. Documentation for the package is given in ACM Algorithm 541 [24]. The original package is written in single precision. In order to get more precise numerical results, a double precision version package is obtained by a little adjustment. The subroutine POIS3D is used to solve for potentials in our model. POIS3D solves the linear system of equations for unknown X values,

```
C1*(X(I-1,J,K) - 2.*X(I,J,K) + X(I+1,J,K)) +
C2*(X(I,J-1,K) - 2.*X(I,J,K) + X(I,J+1,K)) +
A(K)*X(I,J,K-1) + B(K)*X(I,J,K) + C(K)*X(I,J,K+1)
= F(I,J,K),
```

where I = 1, 2, $\cdots$, L, J = 1, 2, $\cdots$, M, and K = 1, 2, $\cdots$,N. The indices K-1 and K+1 are evaluated modulo N, i.e. X(I,J,0)=X(I,J,N) and X(I,J,N+1)=X(I,J,1). The unknowns X(0,J,K), X(L+1,J,K), X(I,0,K), and X(I,M+1,K) are assumed to take on certain prescribed values. The operation count for POIS3D is roughly proportional to

$$L \times M \times N \times (\log_2 M + \log_2 N + 5).$$

REFERENCES

[1] E. J. Adlerman and E. R. Williams, Seasonal variation of the global electrical circuit, *Journal of Geophysical Research*, **101**, 1996, pp.29679–29688.

[2] M. B. Baker, H. J. Christian, and J. Latham A computional study of the relationships linking lightning frequency and other thundercloud parameters, *Quarterly Journal of the Royal Meteorological Society*, **121** 1995, pp.1525–1548.

[3] W. W. Hager, *Applied Numerical Linear Algebra*, Prentice-Hall, Englewood Cliffs, NJ, 1988.

[4] W. W. Hager, Updating the inverse of a matrix, *SIAM Rev.*, **31** 1989, pp.221-239.

[5] W. W. Hager, A discrete Model for the Lightning Discharge, *Journal of Computational Physics*, **144** 1998, pp.137–150.

[6] W. W. Hager, J. S. Nisbet, and J. R, Kasha, The evolution and discharge of electric fields within a thunderstorm, *Journal of Computational Physics*, **82**, 1989, pp.193–217.

[7] W. W. Hager, J. S. Nisbet, J. R, Kasha, and W.-C. Shann, Simulations of electric fields within a thunderstorm, *Journal of the Atmospheric Sciences*, **46** 1989, pp.3542–3558.

[8] L. C. Hale Middle atmospheric electrical structure, dynamics, and coupling, *Earths Middle Atmosphere Advances in Space Research*. **4** 1984. pp.175–186.

[9] J. H. Helsdon, Jr. and R. D. Farley, A numerical modeling study of a Montana thunderstorm: 1. Model results versus observations involving nonelectrical aspects, *Journal of Geophysical Research*, **92** 1987, pp.5645–5659.

[10] J. H. Helsdon, Jr. and R. D. Farley, A numerical modeling study of a Montana thunderstorm: 1. Model results versus observations involving electrical aspects, *Journal of Geophysical Research*, **92** 1987, pp.5661–5675.

[11] J. H. Helsdon, Jr., R. D. Farley, and G. Wu, Lightning parameterization in a storm electrification model, *1988 Proceedings International Conference on Atmospheric Electricity*, pp.849–854

[12] J. H. Helsdon, Jr., G. Wu, and R. D. Farley, An intercloud lightning parameterization scheme for a storm electrification model, *Journal of Geophysical Research*, **97** 1992, pp.5865–5884.

[13] H. W. Kasemir, A contribution to the electrostatic theory of a lightning discharge, *Journal of Geophysical Research*, **65** 1960, pp.1873–1878.

[14] D. R. MacGorman, A. A. Few, and T. L. Teer, Layered lightning activity, *Journal of Geophysical Research*, **86** 1981, pp.9900–9910.

[15] D. R. MacGorman, J. M. Straka, and C. L. Ziegler, A lightning parameterization for numerical cloud models, *Journal of Applied Meteorology*, **40** 2001, pp.459–478.

[16] E. R. Mansell, D. R. MacGorman, Ziegler, and J. M. Straka, Simulated three-dimensional branched lightning in a numerical thunderstorm model, *Journal of Geophysical Research*, **107** 2002, doi:10.1029/2000JD000244.

[17] L. Niemeyer, L. Pietronero, H. J. Wiesmann, Fractal dimension of dielectric breakdown, *Physical Review Letter*, **52** 1984, pp.1033–1036.

[18] F. Pockels, Über das Magnetische Verhalten Einger Basaltischer Gesteien, *Ann. Phys. Chem.*, **63** 1897, pp.195–201.

[19] F. Pockels, Bestimmung Maximaler Entladungs-Strom-Stärken aus Ihrer Magnetisirenden Wirkung, *Ann. Phys. Chem.*, **65** 1898, pp.458–475.

[20] F. Pockels, Über die Blizentlandungen Erreicht Stromstärke, *Phys. Z.*, **2** 1900, pp.306–307

[21] V. A. Rakov and M. A. Uman, *Lightning: physics and effects*, Cambridge University Press, Cambridge, New York, 2003.

[22] F. Rawlins, A numerical study of thunderstorm electrification using a three dimensional model incorporating the ice phase, *Quarterly Journal of the Royal Meteorological Society*, **108** 1982, pp.779–800.

[23] J. C. Strikwerda, *Finite Difference Schemes and Partial Difference Equations*, Champman & Hall, New York, NY, 1989.

[24] P. N. Swarztrauber and R. A. Sweet, Algorithm 541: Efficient Fortran subprograms for the solution of elliptic partial differential equations, *ACM Trans. Math. Software*, **5** 1979, pp.352–364.

[25] T. Takahashi, Determination of lightning origins in a thunderstorm model, Journal of the Meteorological Society of Japan, **65** 1987, pp.777–794.

[26] M. A. Uman, *The Lightning Discharge*, Academic Press, Orlando, Florida, 1987.

[27] United States Air Force Cambridge Research Laboratories, *Handbook of Geophysics*, Macmillan, New York, 1960 (chapter on atmospheric electricity revised in 1970).

[28] R. S. Varga, *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1962.

[29] H. J. Wiesmann and H. R. Zeller, A fractal model of dielectric breakdown and prebreakdown in solid dielectrics, *Jounral of Applied Physics*, **60** 1986, pp.1770–1773.

[30] E. R. Williams, C. M. Cooke, and K. A. Wright, Electrical discharge propagation in and around space charge clouds, *Journal of Geophysical Research*, **90** 1985, pp.6054–6070.

[31] C. T. R. Wilson, On Some Determinations of the Sign and Magnitude of Electric Discharges in Lightning Flashes, *Proceedings of the Royal Society of London Series A*, **92** 1916, pp.555–574.

[32] C. T. R. Wilson, Investigations on Lightning Discharges and on the Electric Field of Thunderstorm, *Philosophical Transactions of the Royal Society of London Series A*, **221** 1916, pp.73–115.

[33] C. L. Ziegler and D. R. MacGorman, Observed lightning morphology relative to modeled space charge and electric field distributions in a tornadic storm, *Journal of the Atmospheric Sciences*, **51** 1994, pp.833–851.

[34] C. L. Ziegler, D. R. MacGorman, J. E. Dye, and P. S. Ray, A model evaluation of non inductive graupel-ice charging in the early electrification of a mountain thunderstorm, *Journal of Geophysical Research*, **96** 1991, pp.833–851.

[35] C. L. Ziegler, J. M. Straka, and D. R. MacGorman, Simulation of early electrification in a three-dimensional dynamic cloud model, *11th International Conference on Atmospheric Electricity*, Commission on Atmospheric Electricity, 1999, pp.359–362.

BIOGRAPHICAL SKETCH

Shu-Jen Huang was born in Taiwan in 1970. She was awarded a Bachelor of Science degree in applied mathematics in 1992 from Fu-Jen University, Taipei, Taiwan. She worked as a teaching assistant at Fu-Jen University for 3 years after her graduation. In 1995, Shu-Jen started her graudate study in mathematics at the University of Florida, from which she received her Ph.D. in mathematics in 2005.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

William W. Hager, Chair
Professor of Mathematics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Jay Gopalakrishnan
Assistant Professor of Mathematics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Shari Moskow
Associate Professor of Mathematics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Sergei S. Pilyugin
Assistant Professor of Mathematics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Vladimir A. Rakov
Professor of Electrical and Computer Engineering

This dissertation was submitted to the Graduate Faculty of the Department of Mathematics in the College of Liberal Arts and Sciences and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

May 2005

_____

Dean, Graduate School

LD
1780
2005
.H874

UNIVERSITY OF FLORIDA
3 1262 08554 2586